

TEXT STRUCTURE INSTRUCTION IN DUTCH PRIMARY EDUCATION

**BUILDING BRIDGES BETWEEN
RESEARCH AND PRACTICE**

Suzanne Bogaerds-Hazenberg

Text structure instruction in Dutch primary education:
Building bridges between research and practice

Suzanne Bogaerds-Hazenberg

ISBN 978-94-6483-085-9. Printed by Ridderprint. Design: Abne Herrebout.

This work was funded by the Research Talent Grant (Grant 406-16-052) of the Dutch Organization for Scientific Research (NWO) and carried out in the context of the ICO research school (Interuniversity Center for Educational Research).

© 2023 Suzanne Bogaerds-Hazenberg. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, without the prior permission from the proprietor.

Text structure instruction in Dutch primary education:
Building bridges between research and practice

Tekststructuuronderwijs op de basisschool:
Bruggen bouwen tussen onderzoek en praktijk
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus
prof. dr. H.R.B.M. Kummeling
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 5 juni 2023 des ochtends te 10.15 uur

door

Suzanne Theresia Maria Bogaerds-Hazenberg

geboren op 13 oktober 1990
te Oldebroek

Promotor

prof. dr. H.H. van den Bergh

Copromotor

dr. J. Evers-Vermeul

Reading committee

prof. dr. H.C.J. de Graaff

dr. E. Mantingh

prof. dr. S.E. McKenney

prof. dr. T.J.M. Sanders

prof. dr. R.C.M. van Steensel

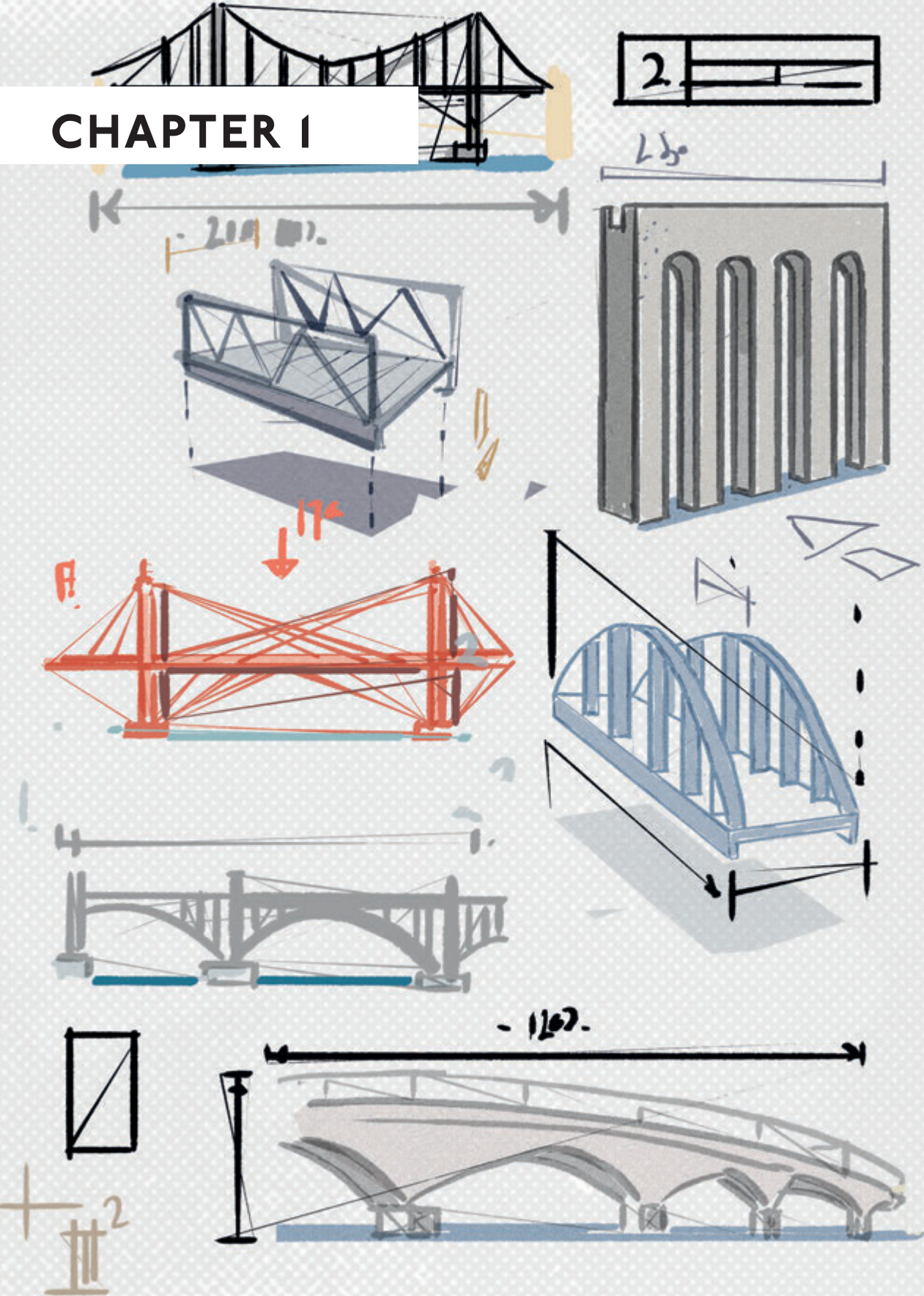
voor Louise, Elodie & Madeleine

Contents

1. Introduction.....	2
1.1 Promoting higher-order comprehension	3
1.2 Teaching text structure: the ecological component.....	8
1.3 Current study and outline	13
2. A meta-analysis on the effects of text structure instruction on reading comprehension in the upper elementary grades	16
2.1 Introduction.....	17
2.2 Theoretical framework.....	20
2.3 Method	29
2.4 Results	43
2.5 Conclusion and discussion	51
3. What textbooks offer and what teachers teach: A mixed-methods analysis of the Dutch reading comprehension curriculum	60
3.1 Introduction.....	61
3.2 Theoretical framework.....	64
3.3 Methods	69
3.4 Results	73
3.5 Conclusion and discussion	90
4. Teachers and researchers as co-designers? A design-based research on reading comprehension instruction	97
4.1 Introduction.....	99
4.2 Theoretical framework: the four design principles	102
4.3 Method	106
4.4 Results	110
4.5 Conclusion and discussion	120

5. Text structure instruction in primary education: Effects on reading, summarization, writing, and metacognitive knowledge.....	126
5.1 Introduction.....	127
5.2 Method.....	131
5.3 Results.....	140
5.4 Conclusion and discussion.....	148
6. A script for success? Teachers' fidelity of implementation of text structure lessons for primary education.....	155
6.1 Introduction.....	157
6.2 Method.....	166
6.3 Results.....	169
6.4 Conclusion and discussion.....	182
7. General discussion.....	192
7.1 Research aim.....	192
7.2 Main findings.....	194
7.3 Discussion of main findings.....	205
7.4 Building bridges between research and practice: recommendations...	216
References.....	220
Appendix A: Meta-analysis.....	256
Appendix B: Mixed-methods analysis.....	261
Appendix C: Example lesson and text structure test.....	262
Appendix D: Model fitting procedures and estimates.....	265
Appendix E: Implementation fidelity measures.....	269
Samenvatting.....	275
Dankwoord.....	290
About the author.....	293

CHAPTER 1



Introduction

Reading comprehension can be defined as ‘the process of simultaneously extracting and constructing meaning through interaction and involvement with text’ (Snow, 2002:11). Reading proficiency is a crucial skill, especially in knowledge-based societies. Therefore, the *European Declaration of the Right to Literacy* states that children should receive ‘evidence-based and cognitively demanding reading and writing instruction that builds both the skill and the desire to read and write’ (ELINET, 2016:9). Yet, Europe faces literacy challenges; one fifth of the students in high school have poor literacy skills (ELINET, 2016). This is problematic, as good reading comprehension skills constitute a major predictor for academic, social, and occupational success (Chall & Jacobs, 1983; Oakhill et al., 2015; Snow, 2002). Reading difficulties can put students at risk for school drop-out (Brasseur-Hock et al., 2011) and can eventually lead to social exclusion (Eloranta et al., 2019).

The high number of students struggling with reading comprehension illustrate the fact that reading comprehension constitutes a very complex skill to acquire (e.g., Cain et al., 2004). Reading difficulties often arise in the primary school years and then tend to persist across adolescence and beyond (Eklund et al., 2015). National and international assessments have shown that a considerable group of Dutch students struggle with reading comprehension, already during primary education (e.g., PIRLS-2016: Gubbels et al., 2017; Mullis et al., 2017; PPON-54: Kuhlemeier et al., 2014; van Weerden & Hiddink, 2013), and in the years beyond (PISA-2015: Feskens et al., 2016; PISA-2018: Gubbels et al., 2019). Even though most students do ultimately attain the required basic levels, there are also concerns about the gradual decline in the percentage of excellent readers (PEIL.Taal-2019; Dutch Inspectorate of Education, 2020; Gubbels et al., 2017).

Various Dutch organizations have expressed the need to improve the quality of reading instruction, and to examine the quality of teaching materials (e.g., Dutch Inspectorate of Education, 2019; Houtveen et al., 2019; Motion Kwint-Peters, 2021; Pereira & Nicolaas, 2019; Raad voor Cultuur & Onderwijsraad, 2019; van den Broek et al., 2021). For example, concerns have been raised that current reading comprehension lessons do not sufficiently foster higher-order comprehension skills, but focus too much on answering questions about texts as a goal in itself

(Berends, 2011; Rooijackers et al., 2020; 2021; Scheltinga et al., 2013; Wieberdink & Kuster, 2015).

This dissertation explores to what extent insights from reading research on higher-order comprehension processes – in particular, on teaching text structures and reading strategies – can be translated into Dutch classroom practice, and explores what is needed to build sustainable bridges between research and practice. In this introductory chapter, I will first introduce the role of reading strategies and text structure knowledge as crucial ingredients to promote students' higher-order comprehension skills (1.1). This is followed by a discussion of three aspects of reading comprehension instruction that deserve extra attention when striving for sustainable changes in the curriculum (1.2). That is, we need a clear view of the role and quality of current teaching materials, gain insight into powerful instructional approaches, and examine teachers' knowledge. A close inspection of teaching materials, instructional approaches, and teachers' knowledge will not only shed light on aspects that can be improved in Dutch educational practice, but also contribute to so-called second-generation text structure research (Williams, 2018). This chapter concludes with the outline of my research project (1.3).

1.1 Promoting higher-order comprehension

Improving the quality of reading comprehension instruction is challenging, especially given the complex and multifaceted nature of reading comprehension in which fluent decoding abilities, vocabulary breadth and depth, prior knowledge, metacognition, and many other factors are involved (Beck et al., 1982; Bus & van IJzendoorn, 1999; Cain et al. 2004; Cromley & Azevedo, 2007; Cutting & Scarborough, 2006; Gough & Tunmer, 1986; Kendeou & van den Broek, 2007; Schoonen et al., 1998; Smith et al., 2021; Soodla et al., 2019; Soto et al., 2019; Stahl & Fairbanks, 1986; van Moort et al., 2020; Verhoeven & Perfetti, 2008).

Most researchers and educators agree that reading instruction in the upper elementary grades should focus on children's higher-order comprehension processes, which entails knowledge of cognitive and metacognitive *reading strategies* (e.g., Bimmel, 2001; Bråten & Anmarkrud, 2013; Brown et al., 1996; Houtveen & van de Grift, 2007; Okkinga et al., 2018b; Rogiers et al., 2019; Soodla

et al., 2017), and knowledge about *text structures* such as compare-contrast or cause-effect (e.g., Duke et al., 2011; Hogan et al., 2011; Meyer & Ray, 2011).

Reading strategies

Reading strategies such as predicting information, questioning, and summarizing are often proposed as important curricular content (Aarnoutse, 2017; Bimmel, 2001; Block & Duffy, 2008; Duke et al., 2011; Palinscar & Brown, 1984; Shanahan et al., 2010), as they can help readers to generate inferences, integrate ideas, solve comprehension problems, and flexibly adapt their reading behavior to the text and reading task at hand (Afflerbach & Cho, 2009; Alexander, 2018; Cain et al., 2004; Cromley & Azevedo, 2007). Generally speaking, high-quality strategy instruction can have positive effects on text comprehension for students in primary education and beyond (Droop et al., 2016; Okkinga et al., 2018a; Palinscar & Brown, 1984; Slavin et al. 2009; Soodla et al., 2017; Spörer et al., 2009), although not all strategy interventions are equally effective (Aarnoutse & Schellings, 2003; Andreassen & Bråten, 2011; Muijselaar et al., 2018; Okkinga et al., 2018b).

Providing reading strategy instruction is quite common in the Netherlands as well (Aarnoutse, 2017; Scheltinga et al., 2003). However, various critical remarks have been made about this practice, as teachers and textbooks seem very preoccupied with answering questions about texts (Rooijackers et al., 2020), and often treat strategies as a goal in itself, rather than a means to an end (Berends, 2011; Houtveen et al., 2019). The usefulness of strategies in certain situations is often not well explained or demonstrated by teachers, which might hinder students' growth in reading comprehension skills (Okkinga et al., 2018b; Wieberdink & Kuster, 2015). Just like a carpenter needs to know when to use screws instead of nails, readers must develop a broad repertoire of strategies and know which strategy to use for a specific text or reading task in order to become self-regulated, strategic readers (Alexander, 2018; Bimmel & Oostdam, 1999; Duke et al., 2011; Malone & Mastropieri, 1992; Rogiers et al., 2019; Wilkinson & Son, 2011).

Reading strategy instruction could become more powerful if it had a clear conditional knowledge component: students should not simply learn which strategies exist (declarative knowledge), or know how to use them (procedural knowledge), but also learn when and why certain strategies are most useful (conditional knowledge; Paris et al., 1983; Winograd & Hare, 1988). Together, these

three types of knowledge can foster a flexible and selective use of reading strategies, which is crucial for deep text comprehension (Alexander, 2018; Rogiers et al., 2019; Simpson & Nist, 2000).

It seems valuable to intertwine reading strategy instruction and instruction about text structures, because the application of reading strategies can be attuned to the text structure (Duke et al., 2011; Hoch & McNally, 2020; Reutzel et al., 2005; Stevens & Vaughn, 2021). For example, the summarization of a cause-effect text requires types of main ideas different from those of a compare-contrast text, and different predictions and visualizations can be made, depending on the text structure. Text structure instruction facilitates students to flexibly and selectively adapt their strategy use to the specific text structure at hand.

Text structure

In most texts, information is organized in a way that suits the rhetorical purpose of the author. For example, information can be organized in a compare-contrast, problem-solution, cause-effect, descriptive, or chronological structure (Meyer, 1975). Various visual and linguistic cues (e.g., signaling words) highlight the structure and make the text cohesive. In Table 1, the text *COVID or influenza? From discovery to pandemic* has a chronological structure signaled by temporal markers. Text structure can be defined as “the organization of ideas, the relationship among the ideas and the vocabulary used to convey meaning to the reader” (Pyle et al., 2017:1; see also Armbruster, 2004).

Text structure is not simply a feature of the text itself, but also affects how readers process textual information. According to the Construction-Integration Model (Kintsch, 1988, 2004, 2013; van Dijk & Kintsch, 1983), readers parse textual input into concepts and relationships, which they need to organize in associative networks. Ideally, readers make connections between ideas within the text, and integrate these ideas with their prior knowledge, until a coherent mental representation of the text arises: the so-called situation model. Text structure can facilitate the construction of a situation model (Bartlett, 1978; Kendeou & van den Broek, 2007; Meyer, 1975; Meyer & Freedle, 1984; Zwaan, 1994). Instead of a default list-strategy that poor comprehenders use, good readers pick up text-

structural cues and use these to organize main ideas accordingly in the situation model (Meyer et al., 1980).

For example, the information on the discovery of COVID-19 in Table 1 should not be read as a list of single facts, but as events in a specific temporal order. Readers who identify the underlying text structure are more likely to read between the lines, find main ideas, and make meaningful inferences about the implicit meaning of the text (Meyer & Ray, 2011; Williams, 2005; Wijekumar et al., 2021), which facilitates deeper text comprehension and recall (Goldman & Rakestraw, 2000; Graesser, 2007; Sánchez & Garcia, 2009; Stine-Morrow et al., 2004; van der Schoot et al., 2010).

TABLE 1
Two text structures and their characteristics with cue words and phrases in italics.

Example	Text structure
COVID <i>or</i> influenza? Influenza and COVID-19 are <i>both</i> contagious respiratory illnesses, <i>but</i> they are caused by <i>different</i> viruses. COVID-19 is caused by a coronavirus, <i>whereas</i> flu is caused by an influenza virus. The symptoms of flu and COVID-19 are <i>similar</i> : <i>both</i> can lead to symptoms like fever and shortness of breath. <i>Compared to</i> flu, COVID-19 seems to spread <i>more easily</i> , and people can be contagious for <i>longer</i> .	Compare-contrast structure The text explains how things are similar and different.
<i>From</i> discovery to pandemic <i>On 1st January 2020</i> , the Wuhan seafood market with wild animals was closed for inspection and disinfection, as it appeared that many visitors of the market developed a new disease. <i>On January 7</i> , Chinese health authorities identified that the disease was caused by a coronavirus that hadn't been encountered in humans <i>before</i> . <i>Five days later</i> the Chinese government shared the genetic sequence of the new virus: COVID-19. <i>On February 7th</i> , the Chinese doctor who warned the world against the virus, died from the disease.	Chronological structure. The text explains the order, cycle, or sequence of events or procedures.

Yet, children often struggle with informational texts: they find it difficult to identify main ideas, and often generate inaccurate inferences, resulting in a shaky and incomplete situation model (Diakidoy et al., 2003; Kim & van Dusen, 1998; Kraal et al., 2018; Meyer et al., 1980; Wijekumar et al., 2020). One of the reasons why informational texts are difficult for children is the fact that the underlying text structures are less familiar to them than the structure of narrative texts, and that different informational text structures require readers to make different types of inferences and connections within and across sentences (Coté et al., 1998; Cromley & Azevedo, 2007; Hiebert & Mesmer, 2013; Sáenz & Fuchs 2002; Schleppegrell, 2004; Wu et al., 2020).

Therefore, it seems important to teach children explicitly about text structures to help them create well-organized situation models. Children can learn how to recognize underlying text structures, and how to use this knowledge to identify and recall main ideas, and to make meaningful inferences (e.g., Meyer et al., 2002; 2011; 2018; Wijekumar et al., 2012; 2013; 2014; 2020; Williams et al., 2004; 2014; 2016; Yuill & Oakhill, 1988). Most empirical research shows that explicit instruction aimed at recognizing text structure has a positive impact on students' reading comprehension (Hebert et al., 2016; Meyer & Ray, 2011; Pyle et al., 2017). It supports students' inferencing in informational texts (Kraal et al., 2018; Stevens & Vaughn, 2021; Wijekumar et al., 2021) and fosters their summarization skills (Miyatsu et al., 2018; Strong, 2020; Westby et al., 2010).

This approach also fits in well with the idea that readers' knowledge and actions to make meaning affect text comprehension, as is emphasized in another influential theory on reading comprehension: the Landscape Model (van den Broek, 2010; van den Broek et al., 1999; 2005). At the heart of this model are the deliberate actions that readers employ to make inferences and to identify important pieces of information (Kucan et al., 2011; van den Broek, 2010). The model explicitly recognizes that readers should know how and when to apply reading strategies to make appropriate inferences and resolve breaks in comprehension.

Text structure interventions often emphasize structure-specific applications of such reading strategies: the way in which readers could summarize or ask questions in a compare-contrast text is different from the way in which they summarize or ask questions in a cause-effect text. Teaching text structure

combined with reading strategies can be a powerful approach to improve children's higher-order reading comprehension (Hoch & McNally, 2020; Read et al., 2008; Reutzel et al., 2005; Stevens & Vaughn, 2021) and foster their metacognitive knowledge (Collins, 1994; Gordon, 1990; Zarrati et al., 2014). This insight forms the basic assumption behind the current study: in order to increase students' text comprehension, we should simultaneously invest in their knowledge about both reading strategies (Landscape Model) and informational text structures (Construction-Integration Model). To put it in other terms: readers should supplement their bottom-up reading strategies with top-down strategies (Stanovich & Stanovich, 1995).

1.2 Teaching text structure: the ecological component

The so-called first-generation text structure research (Williams, 2018) discussed in the previous paragraph has shown that explicit instruction about text structures is a powerful tool to promote students' reading comprehension. However, children's ultimate achievements in reading are also related to the context in which this instruction takes place: curricular materials, instructional approaches, and teacher knowledge (Beerwinkle et al., 2018; Connor et al., 2014; Duke et al., 2011; Odom et al., 2009; Taylor et al., 2003). This instructional context is part of the ecological component in the Component Model of Reading (Aaron et al., 2008), and has recently become more important in second-generation text structure research, which aims to identify variables that modify the effectiveness of text structure instruction, and to improve the quality of the instructional context (Williams, 2018).

Such second-generation research is necessary, as textbook publishers and teachers still struggle to employ evidence-based approaches when it comes to teaching about text structures and reading strategies in informational texts (e.g., Capin et al., 2021; Kucan et al. 2011; Reutzel et al., 2016; Wijekumar et al., 2019a; 2020; 2021). Consequently, this type of research also forms a central theme in this dissertation, paying special attention to the quality of curricular materials, the instructional approach, and teachers' knowledge of text structures.

Curricular materials

First, the quality of teaching materials should be considered as a crucial research topic in the area of reading comprehension (Bohaty et al., 2015; Wijekumar et al., 2021). Teaching materials can be seen as the so-called implemented curriculum (Dockx et al., 2020; Penuel et al., 2014). They form the basis for teachers' classroom enactment and play a prominent role in both what and how students are taught (Aarnoutse, 1990; Aaron et al., 2008; Brown, 2009; Dewitz & Jones, 2013; Valencia et al., 2006; Wijekumar et al., 2019a). The role of teaching materials is even more pronounced when teachers have lower levels of pedagogical content knowledge and rely more on their textbooks (Gudmundsdottir & Shulman, 1987; Valencia et al., 2006).

The content of the reading curriculum has been investigated in various countries (Austria: Seifert, 2021; Belgium: Dockx et al., 2020; China: Zhang et al., 2021; Croatia: Peti-Stantić et al., 2021; Malta: Aguis & Zammit, 2021; Portugal: Cordeiro et al., 2021; US: Beerwinkle et al., 2021). The content and the instructional approach of current Dutch teaching materials remain quite unclear (Hoogeveen, 2018; van den Broek et al. 2021), especially because the constitutional freedom of education principle gives a free hand to educational publishers and schools in the Netherlands to determine their own curricular content and instructional approach (Bruggink & Netten, 2017; Garbe et al., 2016), as only attainment targets are prescribed (Committee Meijerink, 2008).

In other words, although the intended curriculum for Dutch students has been specified, the ways in which teaching materials work towards these targets (the implemented curriculum) and how teachers use these materials in their classroom (the enacted curriculum) remain relatively unclear (see also Motion Kwint-Peters, 2021). Therefore, this dissertation examines the content and pedagogy of textbooks for reading comprehension instruction and teachers' use of these materials in order to unravel the quality of the implemented and enacted curriculum.

Instructional approaches

Second, this dissertation examines various aspects of the instructional approach, as the success of reading instruction depends on the quality of not only the content being taught, but also the instructional approach. Most instructional

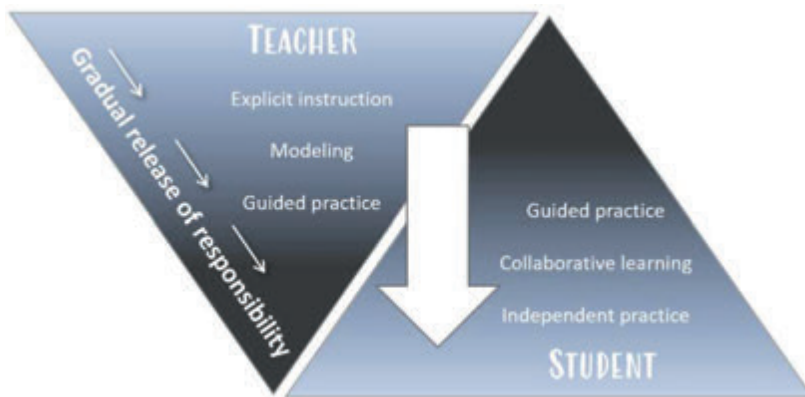
frameworks assume that, after initial teacher-led instruction and modeling (Bissonnette et al., 2010; Duke & Pearson, 2002; Rosenshine & Meister, 1994), the teacher should gradually transfer the responsibility for learning to students with decreasing levels of scaffolding; first through guided practice and teacher feedback (Nolte & Singer, 1985; Pressley et al., 1992), and later on through collaborative small-group learning (Capin et al., 2021; Fuchs et al., 1997; 2021; Klingner & Vaughn 1999; Palinscar & Brown, 1984; Vaughn et al., 2001), until students are capable of practicing on their own. These insights have empirical back-up and are combined in the influential Gradual Release of Responsibility model (GRR model; Fisher & Frey, 2021; Kelly, 2019; Pearson & Gallagher, 1983), which forms the basis of the design and analysis of instructional aspects in this dissertation (Figure 1).

In the Netherlands, current instructional approaches have been debated. It has been criticized that students have to answer comprehension questions about texts – often individually and without a clear functional context – for a large amount of lesson time (Berends, 2011). Not only does this approach overlook the potential of collaborative learning, it might also undermine students' motivation, and limit the transfer of reading comprehension skills to other contexts (Goldman, 1997; Moss, 2004; Patton et al., 2022; van Gelderen & van Schooten, 2011). Moreover, it might not even foster higher-order comprehension skills (Rooijackers et al., 2021).

Yet, little is known about the exact instructional approaches in the Netherlands, although the weekly time spent on reading instruction – probably the simplest indicator of teaching practice – suggests that classroom practices vary widely from school to school (PPON-54; Kuhlemeier et al., 2014). Therefore, I analyzed the way in which the GRR model is applied in Dutch textbooks and classrooms. In addition, I explored whether the GRR model can be used as a guideline for the design of text structure instruction, and what implementation issues arise when teachers follow a GRR model. As little is known about the impact of instructional features on text structure intervention effectiveness (Pyle et al., 2017; Williams, 2018), I also examine in a meta-analysis the moderating impact of instructional features derived from the GRR model on the effectiveness of text structure interventions.

FIGURE 1

Overview of the GRR model: the triangles show the proportion of responsibility divided between teacher and students



Teacher knowledge

Third, effective reading instruction is affected by teacher knowledge, which relates to teachers' ability to successfully enact evidence-based approaches and to compensate for poor textbook content (Beerwinkle et al., 2018; Blömeke et al., 2014; Dewitz & Jones, 2013; Scheerens & Blömeke, 2016; Valencia et al., 2006). Therefore, it is crucial to invest not only in curricular materials, but also in the knowledge of the people who use them (Didion et al., 2020; Duke et al., 2011).

A key concept here is teachers' pedagogical content knowledge (PCK), which refers to teachers' knowledge on how to teach specific subject matter (Gudmundsdottir & Shulman, 1987). It comprises knowledge components such as teachers' subject-specific knowledge on curriculum and assessment, as well as their knowledge on instructional strategies to teach certain content in the best possible way (Janssen et al., 2020; Tuithof et al., 2021). In the area of reading comprehension, PCK entails a good understanding of instructional activities that promote higher-order text comprehension, as well as teachers' ability to analyze texts and identify comprehension problems (Capin et al., 2021; Kucan et al., 2011).

A high level of PCK is important: the more able teachers are to model strategies, explain text structures, and engage in text-based reasoning, the more will they be able to turn their classroom into a powerful learning context. By contrast, a low level of PCK is problematic: it decreases the chance that teachers

provide reading instruction focused on higher-order thinking, which might limit their students' ultimate reading achievements (Kucan et al., 2011; Odom et al., 2009; Okkinga et al., 2018b; Taylor et al., 2003; Wijekumar et al., 2020). For instance, 67% of the teachers in Kucan's study failed to identify main ideas: they treated all ideas as equally important and were often unable to integrate information from different parts of the text. Unfortunately, the teachers "directed their students to do the same" (Kucan et al., 2011:16).

Both nationally and internationally, reasonable concerns about teachers' reading-related PCK have been raised. Many teachers struggle to employ evidence-based approaches, especially when it comes to applying reading strategies in informational texts (Capin et al., 2021; Clark et al., 2017; Hall et al., 2005; Koenig, 2018; Kucan et al., 2011; Piper, 2019; Turcotte et al., 2015; Wijekumar et al., 2019a), or recognizing text structures and main ideas (Beerwinkle et al., 2018; Kucan et al., 2011; Reutzel et al., 2016; Turcotte et al., 2015; Wijekumar et al., 2019a; 2020).

In the Netherlands, concerns on teachers' PCK have been raised as well. For example, the teacher training curriculum hardly prepares future teachers to teach about text structures (Kooiker-den Boer et al., 2019), and Dutch teachers often struggle to successfully model and explain strategies (Okkinga et al., 2018b).

If we acknowledge that teachers' level of PCK is currently inadequate, reading researchers cannot simply develop evidence-based interventions in an ivory tower, drop them in the classroom, and then expect a perfect implementation. Therefore, I explored if design-based research (DBR) – in which primary school teachers and researchers co-design teaching materials – can be a fruitful approach to bridge the research-practice gap. For researchers, DBR creates opportunities to examine and invest in teachers' PCK, as well as to incorporate teachers' practical insights in the newly designed lessons (Broekkamp & van Hout-Wolters, 2007; McKenney & Reeves, 2018; Ormel et al., 2012; Vanderlinde & van Braak, 2010). Involving teachers in a hands-on project under expert guidance might be a powerful way to improve their PCK (see Didion et al., 2020; Hudson et al., 2021).

In addition, I conducted an intervention study and a multiple-case study to explore how well the resulting teaching materials could be enacted by teachers

who were not involved in the lesson design, in order to gain more insight into teachers' needs and knowledge in the area of reading comprehension instruction.

1.3 Current study and outline

The general aim of this dissertation is to gain insight into the effects of explicit text structure instruction on the reading comprehension of Dutch students in primary education. The focus is on grades 4 and 5, because at this age, the development of reading-to-learn skills becomes increasingly important (Chall & Jacobs, 1983) and the exposure to informational texts with complex structures increases (Adams, 2009; Duke, 2000; Schleppegrell, 2004).

Over the past decades, mainly US-based research has shown that text structure can effectively be taught to students of various ages and abilities as a means to improve text recall and comprehension (Burke et al., 2015; Gajria et al., 2007; Hebert et al., 2016; Meyer & Ray, 2011; Pyle et al., 2017). Yet, more research is needed on teaching materials, teacher knowledge, and instructional aspects of text structure interventions (Wijekumar et al., 2021; Williams, 2018), and on the generalizability of text structure instruction to diverse linguistic and cultural contexts (Bohaty et al., 2015). These form important themes in this dissertation.

This dissertation can be divided into three parts. The first part provides the evidence base for text structure instruction in the upper elementary grades. The meta-analysis in *Chapter 2* reveals the most powerful ingredients for effective text structure instruction in grades 4 to 6 by summarizing the findings of 44 high-quality reading interventions focused on text structure. The insights from this meta-analysis are used throughout this dissertation to critically examine current educational practices and to inspire new approaches to teaching reading comprehension.

The second part of this dissertation is focused on identifying the research-practice gap in the area of reading comprehension instruction in Dutch primary education. That is, *Chapter 3* provides a close inspection of current practices: conducting a mixed-methods study, I studied whether and how reading strategies and text structure instruction are currently offered in Dutch textbooks for reading comprehension in grades 4 and 5 (i.e., the implemented curriculum). In addition, this chapter describes teachers' evaluation and use of these materials (i.e., the

enacted curriculum), which also indirectly provides a glimpse into teachers' current level of knowledge with regard to reading comprehension instruction.

The third part of this dissertation examines whether and how the gap between scientific knowledge and current educational practice in reading comprehension instruction in the Netherlands can be reduced. Together with four teachers, I designed and tested a text structure intervention for students in the upper elementary grades: TOP. The lessons were based on four design guidelines derived from findings from the meta-analysis and mixed-methods study (*Chapter 2* and *Chapter 3*). The teachers closely collaborated with two researchers during the whole design process, thereby trying to build sustainable bridges between research and practice. *Chapter 4* reports on this design-based research, with emphasis on the viability of the design principles and the lessons learnt with regard to teacher professionalization.

The resulting lesson series was then experimentally tested in ten classes in a switching-panels experiment, which is presented in *Chapter 5*. I evaluated the effects on various reading outcome measures, summarization, explicit metacognitive knowledge, and writing skills. In order to better understand the outcomes, the multiple case study in *Chapter 6* provides a qualitative examination of teachers' implementation of the intervention.

Chapter 7 forms the closing chapter of this dissertation and summarizes, clarifies and synthesizes the main research findings of the five studies in the light of theoretical and methodological considerations. In addition, implications for practice and future research into reading comprehension instruction are discussed.

Due to the fact that the main chapters in this dissertation have been written as separate papers, some overlap is inevitable. The benefit of this arrangement is that all chapters can be read independently.

CHAPTER 2



A meta-analysis on the effects of text structure instruction on reading comprehension in the upper elementary grades

Abstract

This chapter reports on the results of a meta-analysis in which the results from 44 (quasi-)experimental studies on informative and narrative text structure interventions were synthesized. The focus of the study was to analyze the effects of text structure instruction on reading outcomes for children in grades 4 to 6 in regular school settings. The analysis shows that text structure instruction has positive immediate effects on children's reading comprehension, but that effect sizes vary largely with outcome measure; questions ($g = 0.25$), summarization ($g = 0.57$), recall ($g = 0.37$), and knowledge about text structure ($g = 0.38$). However, students who received text structure instruction no longer outperform control groups at delayed posttests. Content-related features such as a focus on paragraph-level structure, active construction of graphic organizers, and rule-based summarization techniques moderated the effectiveness, but these effects also varied with outcome measure. Instructional features moderated delayed effects: interventions with opportunities for individual student practice resulted in higher delayed effects for comprehension questions. We argue that text structure instruction deserves a place in the primary school curriculum, so that the positive effects on reading can be maintained.

This chapter has been published as:



Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J., & van den Bergh, H.H. (2021). A meta-analysis on the effects of text structure instruction in the upper elementary grades. *Reading Research Quarterly*, 56(3), 435-462. <https://doi.org/10.1002/rrq.311>

2.1 Introduction

Good reading comprehension skills are crucial for understanding text, and play a pivotal role in academic, social and economic success (Oakhill et al., 2015; Rapp et al., 2007). However, reading comprehension is a complex skill, requiring both fluent decoding abilities and good language proficiency (Gough & Tunmer, 1986; Hoover & Gough, 1990; Scarborough, 2001). Both skills need to be promoted through instruction (Bus & van IJzendoorn, 1999; Oakhill et al., 2015). The aim of the current meta-analysis is to examine whether text structure instruction can successfully improve the reading comprehension of students in the upper elementary grades, and to determine what content and instructional components are related to the best outcomes.

According to national standards, by the end of primary education, children should be able to understand simple narrative and informative texts, distinguish various genres, and learn from texts as well (e.g., Common Core Standards Initiative, 2018; Committee Meijerink, 2008). Despite intensive instruction, a substantial number of elementary school students struggle with reading comprehension (Kendeou & van den Broek, 2007). Comprehension problems especially arise when students enter the fourth grade where the development of reading-to-learn skills becomes increasingly important; an effect known as the 'fourth grade slump' (Chall & Jacobs, 1983).

One of the factors that contribute to poor text comprehension is readers' inability to perceive the meaningful relations between information units (e.g., events, facts, settings) in a text (van den Broek & Kremer, 2000). As a result, readers construct a representation of the text base – propositions that are directly derived from the text at the sentence level – but fail to understand how these propositions are organized on a global level. According to the Construction-Integration model (Kintsch, 1988, 2004, 2013; van Dijk & Kintsch, 1983), it is precisely this (re)organization of propositional information at a global level and the successive integration with prior knowledge that are crucial for text comprehension (Stine-Morrow et al., 2004; van der Schoot et al., 2010). The better the information is organized in mental schemata and elaborated with relevant prior knowledge, the more coherent the readers' situation model of the text is, and hence, the better their understanding.

When readers are sensitive to the hierarchical organization of information in texts, this can facilitate the construction of the situation model (Kintsch, 1988; 2004; 2013). Several rhetorical patterns in the organization of information appear in many texts, such as cause-effect, or compare-contrast (Meyer, 1975). Text structure can be defined as “the organization of ideas, the relationship among the ideas and the vocabulary used to convey meaning to the reader” (Armbruster, 2004; Pyle et al., 2017:1). In short, text structure influences how readers read and how writers write (Jiang & Grabe, 2007), and hence not only describes the text itself, but also characterizes readers’ cognitive coherence representations (Meyer & Freedle, 1984; Sanders & Noordman, 2000).

Empirical research has shown that struggling readers typically fail to rely on text structure to guide their reading (Meyer et al., 1980; Rapp et al., 2007; Williams, 2006). By contrast, proficient readers make active use of text structure to organize their memory for textual content; they attend to both the external physical organization of the text (e.g., headings, table of contents) and the internal structure of ideas for a better understanding (Anderson & Pearson, 1984; Kendeou & van den Broek, 2007; Meyer & Rice, 1984). They construct a higher-order structure of text while reading, which ‘guides encoding, recall and reproduction of the text’ (Armbruster et al., 1987:332). Their knowledge about genre or structure influences their expectations about the text, and helps them to better predict and organize textual content while reading (Zwaan, 1994). They are better able to ask relevant questions about the essential text content while reading, which promotes comprehension monitoring (Gersten et al., 2001). From an instructional perspective, explicit teaching of text structure therefore seems a fruitful way to help children better anticipate, predict and monitor their understanding of a text (Ogle & Blachowicz, 2002).

These insights have inspired a wide variety of interventions aimed at teaching the internal structure of texts as a strategy to improve students’ comprehension and recall of texts. Two meta-analyses have shown that text structure instruction is a promising approach for improving text comprehension for learners of various ages (i.e., kindergarten to high school) and abilities (i.e., with and without learning disabilities), with overall effect sizes of $g = 0.56$ (Hebert et al., 2016) and $d = 0.95$ (Pyle et al., 2017). Our meta-analysis refines these meta-analyses by narrowing the focus to students in the upper elementary grades, who are faced with increasing

literacy demands as they transition from 'learning to read' to 'reading to learn' (Chall & Jacobs, 1983).

At the same time, we broaden the scope by including studies on both narrative and informational text structures, and interventions focused on narrow text structure instruction (i.e., recognizing structures). In addition, we included closely related studies addressing structure-based summarization training, paragraph-level structure (e.g., topic sentences) and graphic-organizer instruction. Although these studies do not explicitly train students in naming specific text structures, they still focus students' attention on top-level structures. Moreover, summarization techniques and graphic organizers are often part of text structure interventions, as such activities can promote students' sensitivity to hierarchical discourse patterns in texts, which could facilitate situation-model construction (Kintsch, 1988; 2004; 2013).

Although Pyle et al. (2017) point out that it is important to find out whether and how instructional features (e.g., including collaborative activities) moderate the effectiveness of these interventions, they did not include an analysis of such features. Therefore, our meta-analysis also examines the moderating impact of instructional features such as teacher modeling and collaborative versus individual student practice. This fits in well with recent research suggesting that more attention should be paid to the instructional context of text structure instruction (Beerwinkle et al., 2018; Pyle et al., 2017; Turcotte et al., 2015; Wijekumar et al., 2019a; Williams, 2018).

A growing body of evidence suggests that reading comprehension is not a unitary construct, and that different comprehension tests (e.g., questions, recall, summarization, knowledge) measure different aspects of the reading process (Keenan et al., 2008; Nation & Snowling, 2011). Although previous meta-analyses generated smaller effects on standardized than on researcher-developed measures (Hebert et al., 2016; Pyle et al., 2017), and showed that the largest effects were found on graphic organizer tasks (Pyle et al., 2017), these meta-analyses merged outcome measures when it came to moderator analyses of specific content-related variables. Therefore, it remains unclear whether for instance a focus on paragraph-level structure affects recall, or only summarization.

In terms of maintenance, previous meta-analyses displayed positive delayed effects on reading achievement, although these effects are smaller and typically

lack consistency, and the median delay between immediate and delayed posttest was only seven days (Hebert et al., 2016). However, it is not yet clear how delayed effects are affected by intervention characteristics and if they vary per outcome measure. This is why we examine the impact of content-related and instructional features per outcome measure on immediate and delayed effects, which provides a valuable theoretical and methodological addition. In sum, our meta-analysis evaluates the moderating effects of various content-related and instructional variables in interaction with various outcome measures on students' reading comprehension, in order to refine and expand our knowledge of the ingredients that strengthen text structure instruction.

RQ1: What are the immediate and delayed effects of text structure instruction on text comprehension in grades 4-6, as measured by comprehension questions, recall, summarization, and knowledge of structures?

RQ2: How are these immediate and delayed effect sizes moderated by content-related and instructional features?

The following sections provide an overview of the content-related and instructional features that were taken into account in this meta-analysis, thereby highlighting the state of the art of research on text structure interventions. These features are necessary to describe the categories of analysis and the specific issues that are investigated.

2.2 Theoretical Framework

Content-related features of text structure interventions

Reading instruction in the primary grades often starts with narrative texts, followed by an increasing number of expository texts in the upper elementary grades (Chall & Jacobs, 1983). In the early 1970s, story grammar instruction was developed to aid students' in their comprehension of stories (e.g. Hansen, 1978). However, it was soon discovered that students struggled most with expository text comprehension (Taylor & Beach, 1984), possibly because expository texts contain a high amount of specific vocabulary and many unfamiliar concepts, and

vary more in their underlying text structure than stories typically do (Hiebert & Mesmer, 2013; Pyle et al., 2017). In addition, the reading curriculum at primary schools was strongly focused on narrative texts (Duke, 2000; Durkin, 1978), which resulted in limited exposure to expository texts. As a consequence, students displayed fewer spontaneously developed intuitions about expository text structure (Goldman, 1997).

Over the past decades, many text structure interventions have been developed, which often consist of combinations of the following features:

- Structure recognition (e.g., identifying top-level and paragraph-level structures and/or story grammar);
- Structure visualization (e.g., story mapping, graphic organizers);
- Structure-based summarization (e.g., hierarchical outlining, rule-based summarization).

In the following subsections, we will discuss these features, before we turn to the instructional variables frequently found in text structure interventions.

Structure recognition

Most text structure interventions focus on teaching students how to recognize the top-level structure of expository texts (e.g., compare-contrast, cause-effect). Research on this topic started in the mid-1970s and was strongly influenced by information-processing theories focusing on cognitive processes that affect storage and retrieval of information (Kelly, 2019). One pioneer study was carried out by Bartlett (1978), who found that a training in text structure recognition increased ninth graders' ability to identify the text's top-level structure and use it for recall. Text structure recognition seems to raise text comprehension, especially when multiple text structures are taught (Hebert et al., 2016; Pyle et al., 2017).

Meyer and Ray (2011) provide an excellent overview of interventions focused on structure recognition. Typically, these interventions consist of teaching questions that are answered (e.g., *what are the differences between A and B?*), and practice items in which students categorize short texts as belonging to one structure or another. Also, students learn about cue words or signaling words that frequently appear in these types of structures (e.g., *similar* or *likewise* in compare-

contrast texts), as these words instruct readers how to process an upcoming information segment and how to relate it to a previous one, thereby assisting them toward building coherent text representations (Sanders & Spooren, 2007; van Silfhout et al., 2015). In most interventions, cue words are simply listed in booklets, highlighted or mentioned as characteristics of a specific text structure, but in some interventions students actively highlight (Bohaty, 2005), annotate (Gentry, 2006; Short & Ryan, 1984) or write down cue words (Broer et al., 2002).

Especially in the context of informative texts, structure recognition training can also be focused on the paragraph-level of the text, for instance by teaching students how paragraphs are typically structured in topic sentences, supporting details, and concluding sentences. Often, students receive explicit instruction about the main idea or on topic sentences, and then read a text and select for each paragraph the sentence or phrase that captures the most important information at the highest level (Broer et al., 2002; De Jou & Sperb, 2009; Vidal-Abarca, 1990), or they learn through teacher modeling how to invent a good summarizing phrase when there is no clear topic sentence (Braxton, 2009). Sometimes, students learn more than simply distinguishing between main ideas and details. For instance, Gentry (2006) taught students to make annotations in the margins of paragraphs, like writing *Ex.* when the text discussed an example. In many interventions, paragraph-level instruction was combined with top-level instruction on the features of macro-level text structures.

In other studies, students learned about the blueprint of narrative texts, also called story grammar. This often involves teaching students how to identify the protagonists, their goals or problems, the actions, and the outcome (Gersten et al., 2001; Zwaan, 1994). Students typically receive instruction on the basic elements of a story and use this knowledge to analyze a short story (e.g., Idol & Croll, 1987). Research on these interventions in the upper elementary grades with typically developing students is rather scarce, as many of these interventions are focused on younger children or children with learning difficulties (Gersten et al., 2001). In general, it seems that students with knowledge of story grammar are better able to make predictions about the text, recognize what information is crucial for the plot (Wolman, 1991), and recall more about the main story elements such as 'setting' or 'protagonist' (e.g., Hansen, 1978; Mandler & Johnson, 1977; Weaver & Dickinson, 1982).

In this meta-analysis, we examine the effects of text structure recognition on comprehension, and analyze whether informative and narrative text structure instruction have similar effects, and if the number of structures taught matters.

Structure visualization

Another family of strategies for improving comprehension and recall is to teach children how to visualize the organization of main ideas in graphic organizers. Graphic organizers can be defined as visualizations of the hierarchical relationships of textual information (Griffin et al., 1995), in which relationships between concepts are communicated through the visual placement of concepts relative to each other (Robinson & Molina, 2001). Some common graphic organizers are Venn diagrams, matrices, knowledge maps, and tree diagrams (Manoli & Papadopoulou, 2012), but not outlines and lists as these miss a ‘visual argument’ and are more text-like (Hoffman, 2010).

Graphic organizer development probably started in the early 1970s in Tokyo, where Kaoru Ishikawa developed so-called fishbone diagrams, early cause-effect graphic organizers that were used to control the supply chain and manufacturing process in the ship-building industry (Ishikawa, 1971). In educational contexts, graphic organizers first became in vogue as a variation on advance organizers (Barron, 1980). For a long time, graphic organizers were used in prereading activities to activate and organize students’ prior content knowledge (Moore & Readence, 1984), but nowadays, graphic organizers are also used as a visualization strategy during reading (Leutner et al., 2009), or as post-reading summarization activities. The improved access to digital resources has made computer-based graphic organizers and concept-mapping software (e.g., *Kidspiration* and *Webspiration*) increasingly popular in schools (Ciullo & Reutebeuch, 2013; Smith & Okolo, 2010). Even so, most studies on structure visualization are still centered on print-based modalities.

Graphic organizers constitute a valuable way to enhance comprehension because they can show the main information of the text at a glance and simultaneously clarify relationships between these ideas (Jones et al., 1989). Structure visualization in graphic organizers constitutes a major way to make students aware of text structure, as they provide a “visual map” of the structure (Jiang & Grabe, 2007; Manoli & Papadopoulou, 2012; Pyle et al., 2017). In addition,

graphic organizers can reveal the inferential relationships among text elements (Graesser et al., 1994) and facilitate students' skill in locating specific information (Robinson & Molina, 2002).

Graphic organizers vary in how strictly they map the text structure (Jiang & Grabe, 2007). Some graphic organizers are previously established molds that more closely represent the discourse structure of the text, such as Venn diagrams for compare-contrast texts, or timelines for chronological texts (Ocasio, 2009). These graphic organizers are often instrumentally used in order to directly teach expository text structures (Alvermann & Boothby, 1986; Armbruster et al., 1991; Meyer & Poon, 2001). Typically, students read a text and fill in the empty slots in a partially completed graphic organizer, which afterwards serves as input for a class discussion about text structures (Alvermann & Boothby, 1984; Boothby & Alvermann, 1984; Ermis, 2008; Moore, 1996; Van Steenbrugge, 2006). In some interventions, graphic organizer activities were complemented with writing down signaling words and main ideas (Broer et al., 2002), or students performed writing tasks based on information coming from graphic organizers (Moore, 1996; Raphael et al., 1986).

Other graphic organizers emphasize the hierarchical nature of textual information (e.g., the hierarchical relationships between main ideas and supporting details), without focusing on a specific discourse structure. One example is mapping (Armbruster & Anderson, 1982; Berkowitz, 1986; Griffin et al., 1995) in which main ideas and their relevant relationships are represented in a diagram. For instance, Berkowitz (1986) taught students to write the title in the middle of a sheet of paper surrounded by boxes in which they noted one main idea per paragraph. Interventions were only included in our meta-analysis if the mapping involved representing the discourse structure, or at least the hierarchical relationships between ideas.

In the context of narrative text structure instruction, story maps are used: schematic representations of the key information of narrative texts (Gardill & Jitendra, 1999). A story map can for instance include boxes for setting, goal, plot and outcome (Tackett et al., 1984). Story maps can be used post-reading for summarization, or during reading to help students monitor comprehension and/or highlight main events while reading (Gardill & Jitendra, 1999; Gersten et al., 2001; Idol & Croll, 1987; Tackett et al., 1984). Story grammar can also help

students formulate (Short & Ryan, 1984) or answer questions during reading that help students identify the main constituents of the story (Gordon & Pearson, 1983). As story maps also represent the structure of narrative texts, they were also included in this meta-analysis.

So far, graphic-organizer research has generated mixed results: some studies reveal positive results (Broer et al., 2002; Hoffman, 2010; Ulper & Akkok, 2010; Wijekumar et al., 2012; Wijekumar et al., 2014), other interventions are less effective (Alvermann & Boothby, 1984; Raphael et al., 1986; Walker, 1991), or show that students need a great deal of instructional support to actually benefit from graphic organizers (Griffin et al., 1995). One major issue concerns whether students benefit more from exposure to author-constructed or teacher-constructed organizers than from self-constructed organizers. Some state that simple exposure to graphic organizers may not be sufficient; rather, students may need 'extended instructional training' and 'practice with graphic organizers' before they are able to recognize text structure and make use of this knowledge while reading (Jiang & Grabe, 2007:47). Various studies show, for instance, that the active involvement of students in constructing GOs – even when the graph is already partially complete – facilitates reading comprehension (Berkowitz, 1986; Spiegel & Barufaldi, 1994; Van Steenbrugge, 2006). However, Stull and Mayer (2007) have argued that author-provided organizers are more effective, as student construction of graphic organizers might create cognitive overload.

In this meta-analysis, we therefore examine the effects of structure-based visualizations on comprehension, and analyze whether active construction of structure-based graphic organizers has a positive or negative additional impact on comprehension.

Structure-based summarization

In order to get the gist of a text, readers must overcome the limitations of working memory by ignoring extraneous or redundant information, and focusing specifically on macro-level information, such as topic sentences (Bean & Steenwyk, 1984; Kintsch & van Dijk, 1978). This process of eliminating and reworking information can be promoted through summarization instruction (Armbruster et al., 1987; Bean & Steenwyk, 1984; Elledge, 2013; Frey et al., 2003; Taylor, 1986; Westby et al., 2010).

Teaching summarization improves both the quality of written summaries and students' overall text comprehension (Duke et al., 2011; Taylor & Beach, 1984). A recent literature review states that the defining characteristic of successful summarization training is the emphasis on main idea identification and text structure recognition (Miyatsu et al., 2018). In fact, text structure can scaffold students' summarization skills, as it provides them with tools and heuristics to distinguish main ideas from unimportant information (Hogan et al., 2011; Meyer et al., 1980; Stevens, 2018; Taylor, 1985; Winograd, 1984). In addition, it helps students understand how these main ideas are organized, which helps them to write coherent summaries (Miyatsu et al., 2018). Structure-based summarization strategies might also facilitate text recall, as the text structure might function as a mnemonic aid (Taylor, 1982).

In most elementary schools, teachers refrain from providing explicit instruction about specific summarization techniques (Beerwinkle et al., 2018; Elledge, 2013; McKeown et al., 2009), or struggle themselves with the identification of main ideas (Kucan et al., 2011; Turcotte et al., 2015; Wijekumar et al., 2020). It is therefore not remarkable that many students struggle to identify main ideas (Baumann, 1983), seldom formulate summarizing topic sentences (Garner, 1983), and use deletion of propositional expression as their main summarization strategy (Winograd, 1984). Over the past decades, researchers have come up with various summarization techniques that are less intuitive, but instead rely on the external or internal structure of texts.

The hierarchical summarization strategy (e.g., Taylor, 1982; 1985) consists of skimming the *external* organizational text structure first – the headings and subheadings – and then prepare a skeletal outline based on these headings. Next, students write down one main idea per section. Yet, this strategy might be limited to texts with an unambiguous heading-subheading format (Armbruster et al., 1987).

The rule-based summarization strategy (Brown et al., 1983; McNeil & Donant, 1982) relies more heavily on the *internal* structure of paragraphs (i.e., identifying topic sentences). This strategy provides students with a set of six summarization rules, based on the work of Kintsch and van Dijk (1978), such as 'delete redundant information' or 'invent a topic sentence'. These rules help students to first eliminate information and then rework the remaining bits into a coherent

summary (Brown et al., 1983; McNeil & Donant, 1982). Students can learn to highlight topic sentences, to circle words that must be replaced by superordinate concepts and cross out trivia (Braxton, 2009). Rule-based summarization is often combined with text structure recognition. For instance, students learn how signal words and structure-specific questions can be used to identify main ideas (e.g., Elledge, 2013).

Not only the internal structure of the paragraph, but also that of the whole text can function as a framework for summarization (Armbruster et al., 1987; Inchausti et al., 2009; Ocasio, 2006; Stevens, 2018; Vidal-Abarca, 1990). For instance, in the studies by Armbruster et al. (1987) and Stevens (2018), students were taught about the characteristics of the problem-solution text structure. Then, students received a specific problem-solution frame in which they could summarize the main point of a text, as well as a list of structure-specific guidelines for their summary (e.g., *'Sentence 1: Tells who had a problem and what the problem is'*). Vidal-Abarca (1990) and De Jou and Sperb (2009) explained various text structures and modeled where and how to find the main idea in these texts.

In this meta-analysis, we examine the effects of structure-based summarization techniques, and analyze whether a rule-based summarization approach with its emphasis on internal text structure has an additional impact on text comprehension.

Instructional features of text structure interventions

Text structure instruction can be provided in many different ways. Previous meta-analyses have shown that the implementer plays a crucial role: researcher-taught interventions are often more effective than teacher-taught interventions (e.g., Dignath & Buettner, 2008), also when it comes to text structure instruction (Pyle et al., 2017). Recently, more attention has been paid to the so-called ecological component of text structure instruction, by investigating how teachers explain text structure and other evidence-based strategies in their classroom (Beerwinkle et al., 2018; Wijekumar et al., 2019a), and how teachers can be professionalized in teaching text structures (Reutzel et al., 2016; Wijekumar et al., 2019b).

Meyer and Ray (2011) emphasize that teachers should have access to adequate instructional materials for modeling and direct instruction, for instance by providing them with "intelligent tutors or scripted lessons" (p. 138). Williams

(2018) also makes a plea for second-generation text structure research that goes beyond developing excellent materials, and focuses more on the “context in which the instruction occurs” (p.10). In addition, the previous meta-analyses state that future research should examine the mediating role of instructional features, as both the implementer and the type of instructional activities might affect the effectiveness of text structure interventions (Hebert et al., 2016; Pyle et al., 2017).

The *Gradual Release of Responsibility Model* (Fisher & Frey, 2021; Pearson & Gallagher, 1983) gives a useful framework for describing and comparing current teaching practices. This model suggests to sequence various instructional activities such that the responsibility for the learning process relies first mainly on the teacher (e.g., direct instruction, modeling) and is then gradually transferred to the student with decreasing levels of scaffolding (e.g., guided practice, collaborative activities, individual activities). This also reflects the idea that reading comprehension lessons should follow a pattern of stepwise phasing out the teacher, while phasing in the students. They gradually take over the lead from teachers by applying comprehension strategies first in small groups, then in pairs, and finally individually (Nolte & Singer, 1985; Singer & Donlan, 1980). Many studies on reading instruction have emphasized the importance of direct instruction of cognitive and metacognitive strategies (e.g., Houtveen & van de Grift, 2007; Rosenshine & Meister, 1994; Soodla et al., 2017; Stevens et al., 1991).

Research suggests that modeling plays a pivotal role in increasing reading comprehension, especially when the demonstration of the model includes conditional knowledge: why the model is doing something, as well as metacognitive and motivational aspects (Kostons et al., 2014). Modeling can raise students' self-efficacy to carry out tasks on their own (van Gog & Rummel, 2010), which may be particularly beneficial to students with low reading self-efficacy. In various studies, positive effects have been found for explicit instruction combined with teacher modeling (e.g., Collins & Pressley, 2002).

For reading and many other areas of instruction, collaboration has been a successful way to enhance learning. If well implemented, it can improve students' time on task (Cohen & Benton, 1988), raise academic performance (Slavin, 1987), and increase the quantity and quality of student interactions, to mention a few benefits (Fuchs et al., 1997; Garibaldi, 1979; Johnson & Johnson, 1984; 2017; Vaughn et al., 2001).

Text structure interventions in the upper elementary grades show variability in terms of their instructional approach. Most interventions are composed of direct instruction (with or without modeling) followed by individual practice (e.g., Armbruster et al., 1987; Berkowitz, 1986; Bowman & Gambrell, 1981; Elledge, 2013; Gordon & Pearson, 1983; Griffin et al., 1995), whereas in other studies, students only work in groups (e.g., Darch et al., 1986), or both in groups and individually (e.g., Fitzgerald & Spiegel, 1983; Moore, 1996; Scott, 2011; Tackett et al., 1984; Van Steenbrugge, 2006).

Only a handful of studies display the full gradual release of responsibility with explicit instruction, modeling, collaborative and individual work (e.g., Braxton, 2009; Ermis, 2008; Gentry, 2006). Recent work has focused on the development of web-based text structure teaching such as ITSS; Intelligent Tutoring of the Structure Strategy (e.g., Meyer et al., 2010; Wijekumar et al., 2012; 2013; 2014), and on the effectiveness of specific instructional features, such as the presence or absence of tutors (Meyer et al., 2002), or instruction tailored to students' individual needs (Meyer et al., 2010).

Except for analyzing the effects of online tutors versus teachers (Pyle et al., 2017), prior meta-analyses have not yet analyzed the effects of instructional components on the effectiveness of text structure instruction. Therefore, our meta-analysis examines whether it matters if teacher and student activities follow a pattern of gradual release of responsibility.

2.3 Method

Inclusion and exclusion criteria

A set of inclusion and exclusion criteria was developed that guided the iterative search and selection procedure. Studies were deemed eligible for inclusion in the meta-analysis if the following five criteria were met:

1. The study was published in English, German, French, Dutch, or Spanish between 1974 and 2018, and was available online or could be retrieved directly from the authors.

2. The study focused on students in general primary education in grades 4-6. Participants could be children with mild reading difficulties, but not children with severe learning or reading difficulties, hearing problems, and/or second language learners.

3. The treatment group was taught about informative text structures (such as *description*, *compare-contrast*, *problem-solution*, *cause-effect* and *chronology/sequence*; e.g., Meyer, 1975) or story structure. Interventions in which text structure was explicitly used as part of a summarization or visualization technique were also included.

4. The treatment group was compared to a business-as-usual control group or to a control group receiving an alternative instruction. There was no restriction on sample size or sampling procedures, as long as procedures were well documented.

5. The study included at least one posttest focused on the comprehension of texts. The posttest could consist of one or several (non-)standardized comprehension question tests, cued or free recalls, graphic organizer completion and/or summarization tasks and text structure knowledge tests.

Studies were excluded when the researcher(s) did not provide the statistics necessary to calculate a weighted effect size, or when results were summarized over multiple age groups, such that it was impossible to calculate effect sizes for grades 4 to 6 exclusively, even after contacting authors. Correlational and qualitative studies were excluded, as well as studies with a within-subjects design or a multiple-baseline design (e.g., Haria & Midgette, 2014).

Search and selection procedure

A four-step process was used to conduct a comprehensive search. First, studies were located by using the electronic databases of *ERIC*, *Web of Science*, *PsychInfo* and *Google Scholar*. Second, databases of theses and dissertations (*PQDT*, *Ethos*, *Open Thesis*) were searched to locate unpublished studies. Third, a series of Dutch, French, German and Spanish search engines was used to identify relevant studies in languages other than English. Fourth, we conducted a cited reference search of previous reviews and meta-analyses (Hebert et al., 2016; Jiang & Grabe, 2007; Meyer & Ray, 2011; Pyle et al., 2017) and checked reference lists of the studies that were judged eligible.

In the literature search, we used the keywords *text structure* or *top-level structure* and *reading (comprehension)* combined with age group indicators (*primary, elementary, fourth/fifth/sixth grade*). In a second search, we included keywords specifying different text structures (*description, enumeration, classification, cause effect, compare contrast, sequence, chronology, problem solution, story (grammar), narrative*) or keywords on the potential ingredients in text structure interventions (*topic sentence, signaling words, cue words, graphic organizer, schematizing, main idea, outline, summary*) in order to maximize the number of articles located. This search yielded approximately 2900 results.

After removal of the many duplicates from the list and a first title screening, 408 abstracts were imported for closer examination. Abstracts and method sections were read to determine whether these studies qualified for inclusion. We removed 355 articles, because participants did not match the required age and aptitude profile ($n = 78$), or because studies did not meet the criteria for text structure instruction ($n = 208$; e.g., reciprocal teaching interventions or semantic mapping studies without a focus on text structure), were focused on text manipulations ($n = 45$), did not measure text comprehension ($n = 20$), or were published in another language ($n = 4$), which left us with 52 eligible studies. Eight studies could not be located, or provided insufficient statistics for effect size calculation, such that the final set consisted of 44 studies published between 1982 and 2018.

Study feature coding

Studies were coded on publication type, participants (number, grade), research design (quasi-experimental vs. experimental), type of outcome measure (comprehension questions, free recall tasks, summarization tasks, knowledge tests), test type (standardized vs. non-standardized), measurement occasion (immediate vs. delayed), and reference group (business as usual vs. alternative intervention as control). We also coded text features, such as genre (narrative and/or informative), the number and exact type of structures (e.g., narrative, description, cause-effect, compare-contrast, problem-solution, chronology). In addition, we coded various content-related and instructional features of both the treatment and control condition (Table 1).

The instructional content of the intervention was described by using three higher-level descriptors that formed the basis for further analysis. These descriptors were not mutually exclusive: one intervention could for instance include both training in text structure recognition and structure visualization. To refine these categories, we also added variables describing whether there was an additional focus on paragraph-level structure, active construction of graphic organizers, and/or instruction in the rule-based summarization technique.

Two instructional variables were defined: teacher activities and student activities. Instruction activities could be primarily focused on telling (i.e. explicit instruction) or showing (i.e. modeling). Student activities were either individual or collaborative. The analysis of these instructional components was based on the description of the intervention and procedures in the article, or on an examination of the examples of materials that were provided (e.g., screenshots, examples, teaching materials in appendices). If the text mentioned that teachers used 'think-aloud protocols' or 'demonstrated the strategy', this was coded as modeling. Similarly, if the authors mentioned 'working in pairs' or 'small group activities', we scored this as a collaborative activity. If nothing was mentioned or could be derived logically, this was coded as absence of the instructional feature at hand.

Apart from the distinction between standardized and non-standardized tests, we also distinguished between four types of outcome measures: comprehension questions, free recall tasks, summarization tasks, and knowledge tests. Outcome measures coded as comprehension questions were all tests that

involved multiple-choice items and short-answer questions with the text present (e.g., literal comprehension, referential comprehension, interpretation questions).

Recall tasks included tasks in which students were typically asked to read a text, and then performed a memory task without the text present. This could be writing or telling everything they remembered or only the most important information, or cued recall tasks in which they answered factual questions about text content (e.g., where did the story take place?). Summarization tasks included all tasks in which students had to summarize part of a text with the text present, such as student-generated written summaries, but also graphic organizer completion, and tasks focused on highlighting the main ideas. Text structure knowledge tasks included all tests focused on knowledge about specific text structures, such as tasks in which students had to identify the structure of a text segment (e.g., Stevens, 2018), match segments to a text with a similar structure (e.g., Broer et al., 2002), or select the right structure-specific signaling words in cloze tasks (e.g., Wijekumar et al., 2013).

TABLE 1
Description of intervention content and instructional components

Descriptor	Description
Structure recognition and/or focus on the paragraph level	Students are explicitly taught about the internal structure of texts or paragraphs and practice with recognition. The focus on the paragraph level might entail instruction about topic sentences and supporting and concluding sentences.
Structure visualization and/or active construction	Students receive instruction on schematic representations of text structure and content. They study and/or fill out or actively create maps and graphic organizers.
Structure-based or rule-based summarization	Students learn how to summarize a text on the basis of headings and other hierarchical outlining principles, or learn

techniques for paraphrasing main ideas or condensing text by strictly following a set of rules.

Teacher-led instruction ± modeling	The teacher provides explicit instruction and/or demonstrates structure recognition, structure visualization, or structure-based summarization techniques by thinking aloud in front of the class.
Student activities ± individual practice	Students have opportunities to practice their skills, collaboratively and/or individually.

All studies were coded by the first author, and a random sample of four studies (10% of the total sample) were coded by a trained research assistant, with 94% total interobserver agreement and kappa-values ranging from $\kappa = .83$ (instructional approach) to $\kappa = .95$ (methodological descriptors). Table 2 provides an overview of all studies and shows immediate and delayed effect sizes per outcome measure.

Effect size calculation and statistical analyses

For calculation of effect sizes, we used Hedges’ *g* which is almost similar to Cohen’s *d* (Cohen, 1988; Fritz et al., 2012), but provides an unbiased estimate of effect sizes for the few studies with smaller sample sizes ($n < 20$) that were included in this meta-analysis (Borenstein et al., 2011; Cumming, 2012). Although effect sizes of around .20 are generally classified as small, we interpret them as meaningful, because they were obtained in educational contexts, where even effect sizes that are generally classified as small, are of interest (Durlak, 2009; Hedges & Hedberg, 2007). Moreover, the effect sizes we report in this meta-analysis should be interpreted as *additive effects*, as they represent what students in a text structure condition gain extra on top of what students learn in an alternative intervention or the regular reading curriculum.

We calculated the standardized mean differences (Hedges’ *g*), for immediate and delayed effects separately. For the calculation of immediate effects, we subtracted the mean difference in performance of the control group (immediate posttest-pretest) from the mean difference in performance of the treatment

group (immediate posttest-pretest), divided by the pooled standard deviation of the two groups.

As we were interested in long-term differences between experimental and control groups as well, we also calculated delayed effect sizes. These are indicative of the effects of text structure instruction that remain, over and above what students learn in business-as-usual. This calculation was based on comparing the mean performance of the control group to the treatment group on delayed posttests, divided by the pooled standard deviation of the two groups.

Therefore, a delayed effect of 0 would mean that there are no lasting differences between experimental group and control group, whereas a delayed effect of +1 would indicate that the experimental group outperforms the control group by one standard deviation at delayed posttests, either because the experimental group made more progress than the control group, or because the control group performed worse, while the experimental group maintained their skill compared to the immediate posttest.

For six studies that did not report the exact means and standard deviations, we calculated effect sizes based on Analyses of Variance (ANOVAs). As the sample sizes of these six studies were quite large ($n > 42$), the calculated value of Cohen's d was practically identical to Hedges g (Borenstein et al., 2009; Fritz et al., 2012).

When multiple outcome measures were used, separate effect sizes were calculated per type of outcome measure (questions, recall, summarization, and text structure knowledge), so that we were able to show the impact of the intervention variables on each measure separately. Also, separate effect sizes were calculated when multiple text structure interventions were compared within studies. For instance, Ulper and Akkok (2012) investigated the effectiveness of structure-based summarization strategies, as well as the effectiveness of this approach in combination with training in text structure recognition. In these instances we calculated separate effect sizes per condition, even though these are presented as averaged effect sizes in Table 2.

TABLE 2

Overview of all 44 studies included in the meta-analysis

Study	Publi- cation type ^a	Grade	N ^b	Intervention(s) vs. control	Genre ^c	N of ESs	Effect sizes (Hedges' g)			
							Questions	Recall task	Sum- mary	Text structure tests
Studies primarily focused on text structure recognition (n = 20)										
Bohaty (2015)	D	4 & 5	45 L	Text structure recognition and focus on cue words vs. BAU	CC, CE, D, PS, CH	3	-0.17 ^d	0.13		0.99
Carriedo & Alonso- Tapia (1996)	J	6	211	TSL with a strong focus on main idea identification vs. traditional questioning	CC, D, N, P, PS	2			0.58	-0.24
Fitzgerald & Spiegel (1983)	J	4	19	Story grammar vs. vocabulary and dictionary instruction	N	6	-0.13	-0.02		
Gentry (2006)	D	4	30	Recognizing structures and with text annotation vs. vocabulary instruction	CC, CE, D, PS, CH	2	0.11 vs. -0.08 ^d			

Study	Publi- cation type ^a	Grade	N ^b	Intervention(s) vs. control	Genre ^c	N of ESS	Effect sizes (Hedges' <i>g</i>)			
							Questions	Recall task	Sum- mary	Text structure tests
Gordon & Pearson (1983)	R	5	42	Story grammar vs. inference awareness	N	6	0.29 vs. 0.43 ^d	0.69		
McDermott (1991)	D	4	66	Recognizing structures vs. BAU	D, PS	2		0.66		
Meyer et al. (2002)	J	5	60	Web-based structure strategy (± tutoring) vs. BAU	CC, CE, D, PS, CH	16	0.19 vs. 0.61 ^d	0.41 vs. 0.05 ^d	0.30 vs. 0.15 ^d	0.50 vs. 0.43 ^d
Meyer et al. (2011)	J	5	131	Web-based structure strategy (individualized) vs. web-based structure strategy (standardized)	CC, PS	12	0.42	0.08	−0.07	0.45
Ponce et al. (2012)	J	4	104	Recognizing structures (online) vs. BAU	CC, CE, P, CH	1	0.71			
Raphael et al. (1986)	R	5 & 6	159	TSl (writing ± social context) vs. BAU	CC, D, N, PS	2	0.06		0.38	
Raphael & Kirschner (1985)	R	5 & 6	45	TSl with writing vs. BAU with writing	CC	2		1.22	0.82	
Scott (2011)	D	6	205	TSl vs. content-focused instruction (social studies)	D, CH	4	−0.27		0.95	

Study	Publi- cation type ^a	Grade	N ^b	Intervention(s) vs. control	Genre ^c	N of ESs	Effect sizes (Hedges' <i>g</i>)			
							Questions	Recall task	Sum- mary	Text structure tests
Short & Ryan (1984)	J	4	56	Story grammar vs. attribution L training	N	2	0.62	0.66		
Spires et al. (1992)	J	4	74	TSL vs. previewing structure vs. traditional questioning	CC, PS	16	0.40 vs. 0.08 ^d		0.04 -.35 ^d	
Tackett et al. (1984)	J	6	45	Story grammar vs. exposure to L well-formed stories vs. BAU	N	2		0.54		
Troyer (1994)	P	4–6	173	TSL and mental models vs. TSL ± GO vs. BAU	CC, D	14	0.37 vs. 0.20 ^d			0.40
Walker (1995)	J	5	105	TSL vs. traditional questioning	CC, CE, D, PS, CH	2	0.09 vs. -0.15 ^d			
Wijekumar et al. (2012)	J	4	260 0	Web-based structure strategy vs. BAU	CC, CE, D, PS, CH	7	0.10	0.14	0.49	0.20
Wijekumar et al. (2013)	J	4 & 5	443	Web-based structure strategy vs. BAU	CC, CE, D, PS, CH	8	0.19	0.46	0.74	0.39
Wijekumar et al. (2014)	J	5	264 5	Web-based structure strategy vs. BAU	CC, CE, D, PS, CH	7	0.20	0.26	0.53	0.29

Study	Publi- cation type ^a	Grade	N ^b	Intervention(s) vs. control	Genre ^c	N of ESs	Effect sizes (Hedges' <i>g</i>)		
							Questions	Recall task	Sum- mary Text structure tests
<i>Studies primarily focused on text structure visualization (n = 13)</i>									
Alvermann & Boothby (1986)	J	4	24	GO training (14 weeks) vs. GO training (7 weeks) with TSI vs. reading recitation	D	9	0.23	0.43	
Armbruster et al. (1991)	J	4 & 5	365	GO training vs. BAU	E	2	0.11		
Berkowitz (1986)	J	6	99	GO construction vs. GO study vs. questioning vs. rereading	E	40	0.08 vs. −0.05 ^d	0.46 vs. 0.05 ^d	
Boothby & Alvermann (1984)	J	4	38	GO training vs. no specific instruction	CE	5	0.99	0.98 vs. 0.17 ^d	
Bowman & Gambrell (1981)	P	6	100	Story maps and questioning vs. traditional questioning	N	4	0.46 vs. 0.08 ^d	0.19 vs. 0.32 ^d	
Broer et al. (2002)	J	6	354	Recognizing structures with emphasis on schematics vs. traditional questioning	CE, D	7	0.15		0.55 vs. 0.18 ^d
Darch et al. (1986)	J	6	84	GO training (group) vs. directed reading vs. GO training (individual)	E	4	0.71 −0.54 ^d		

Study	Publi- cation type ^a	Grade	N ^b	Intervention(s) vs. control	Genre ^c	N of ESs	Effect sizes (Hedges' <i>g</i>)			
							Questions	Recall task	Sum- mary	Text structure tests
Ermis (2008)	C	4 & 5	50	GO training with some TSI vs. traditional questioning	CE, D	2	0.67			
Griffin et al. (1995)	J	5	99	Explicit vs implicit GO training Explicit vs implicit main idea training	E	24	0.35 -0.25 ^d	0.04 -0.42 ^d		
Hoffman (2010)	D	5	153	GO training and metacognitive monitoring vs. no instruction vs. GO training	CC	4	0.33			
McLaughlin (1990)	D	5	68 L	GO training with TSI vs. other cognitive strategies	CC	3	0.10	0.04	0.21	
Moore (1996)	P	6	76	GO training and TSI vs. traditional questioning	CC, CE	2	0.72 vs. 0.40 ^d			
Van Steen- brugge (2006)	M	5 & 6	103	GO construction vs. GO study vs. BAU	CE, D, N, CH	3	0.05			

Study	Publi- cation type ^a	Grade	N ^b	Intervention(s) vs. control	Genre ^c	N of ESS	Effect sizes (Hedges' <i>g</i>)		
							Questions	Recall task	Sum- mary Text structure tests
<i>Studies primarily focused on structure-based summarization (n = 11)</i>									
Armbruster et al. (1987)	J	5	82	Rule-based summarization and TSl vs. traditional questioning	PS	2	0.64		1.66
Bean & Steenwyk (1984)	J	6	60	Rule-based summarization vs. intuitive summarization vs. no specific instruction		4	0.80		0.74
Braxton (2009)	D	4 & 5	64	Rule-based summarization vs. intuitive summarization	CC, CE, D, PS, CH	4	-0.23		1.82
De Jou & Sperb (2009)	J	5	86	Main idea identification training with some TSl and metacognitive strategies vs. BAU	E, N	4	0.35	0.66	
Elledge (2013)	D	4-6	126	Rule-based summarization and TSl vs. BAU	CE, D	6			0.10
Ocasio (2006)	D	5	63	Structure-based summarization and TSl vs. reciprocal teaching	CC, CE, PS, CH	1			2.36
Stevens (2018)	D	4 & 5	61 L	Main idea identification and TSl	CC, D, PS	3	-0.14	0.70	0.58

Study	Publi- cation type ^a	Grade	N ^b	Intervention(s) vs. control	Genre ^c	N of ESs	Effect sizes (Hedges' <i>g</i>)			
							Questions	Recall task	Sum- mary	Text structure tests
Taylor (1982)	J	5	48	Hierarchical summarization vs. traditional questioning	E	2	-0.07			0.91
Taylor (1985)	J	6	93	Hierarchical summarization ± TSI with writing vs. traditional questioning and writing	CC	18	0.27 vs. -0.47 ^d	0.25 vs. 0.51 ^d	0.39 vs. -0.01 ^d	
Ulper & Akkok (2010)	C	6	13L & 15H	Structure-based summarization ± TSI vs. BAU	CH	2			0.90	
Vidal- Albarca (1990)	J	5	51	Main idea identification and some TSI vs. BAU		3			1.06	1.04

Note. N of ESs = number of effect sizes; GO = graphic organizer; TSI = text structure instruction; BAU = business as usual condition.

^aC = book chapter; D = doctoral dissertation; J = journal article; M = master's thesis; P = paper; R = report.

^bH = high-achieving students; L = low-achieving students. ^cCC = compare-and-contrast; CE = cause-and-effect; D = descriptive; E = expository (no further specification); N = narrative; P = persuasive; PS = problem-and-solution; CH = chronology/sequence. ^dA delayed effect (i.e., immediate posttest-delayed posttest difference).

Because the handling of multiple effect sizes from one study leads to statistical dependencies in the data, we aggregated data sets per outcome measure. By taking this approach, no more than two effect sizes per study were calculated: one delayed ES and one immediate ES. In the analyses, more weight was allocated to studies with larger sample sizes (Borenstein et al., 2011). Mixed-effect model analyses were run per outcome measure, so that we actually conducted four parallel meta-analyses, in which we subsequently added methodological, content-related, and instructional variables as moderators. After constructing full factorial models, we simplified the models with only the relevant parameters so that the additive effect of all parameters could be estimated, without running the risk of overfitting.

Due to differences in methodological, content-related and instructional features, we could not make the assumption of one common effect size. Therefore, random effects models were used, which assume not one true effect size, but an effect size distribution. This made it possible to generalize to populations beyond the included studies (Borenstein et al., 2011). We also examined the within-class goodness-of-fit by conducting homogeneity tests (Cooper, 1998) in order to check whether the variability in effect sizes was so large that moderator analyses were needed. The differences in fit of subsequent (nested) random effects models were tested by means of loglikelihood-ratio tests. All effect size calculations and moderator analyses were conducted in R using the Metafor package (Version 3.3.3; Viechtbauer, 2010). A full overview of our stepwise model fitting can be found in Appendix A.

2.4 Results

For each outcome measure, a random effects model was used to assess the overall average effect size. The overall effect sizes were for questions $g = 0.14$, with a 95% confidence interval (CI) ranging from 0.03 to 0.25.; for recall, $g = 0.30$ [95% CI: 0.19, 0.41], for summarization $g = 0.43$ [95% CI: 0.24, 0.61], and for text structure knowledge $g = 0.34$ [95% CI: 0.28, 0.41]. We have to be careful with the interpretation of these mean effect sizes, as for all outcome measures there was significant heterogeneity in effect sizes ($Q_{questions}(88) = 310.37, p < .001$; $Q_{recall}(84) = 248.38, p < .001$; $Q_{summarization}(51) = 328.47, p < .001$; $Q_{knowledge}(26) = 79.87, p < .001$).

.001). Of course, this heterogeneity can partly be attributed to the fact that the overall effect sizes still include both immediate and delayed effects, and various types of research designs.

For questions and text structure knowledge, we were unable to show a publication bias ($\Delta\chi^2(1) = 1.32, p = .25$ and $\Delta\chi^2(1) = 0.23, p = .63$). However, for recall and summarization we found evidence for a reversed publication bias ($\Delta\chi^2(1) = 5.00, p = .03$ and $\Delta\chi^2(1) = 3.95, p = .047$). That is, effect sizes for recall and summarization reported in peer-reviewed scientific journals were systematically smaller than the unpublished ones.

Methodological variables

Control condition

In some studies ($n = 18$), the authors explicitly mention that the text structure intervention was compared to an alternative instruction instead of to a business-as-usual control group. For instance, text structure instruction was compared to vocabulary instruction (e.g., Fitzgerald & Spiegel, 1983; Gentry, 2006), instruction in cognitive reading strategies such as making predictions and inferences, and activating prior knowledge (e.g., Gordon & Pearson, 1983; McLaughlin, 1990; Ocasio, 2006), or more intuitive summarization strategies (Bean & Steenwyk, 1984; Braxton, 2009).

However, in most studies, the control groups continued to follow their usual reading curriculum during the intervention. This business-as-usual typically involved a traditional approach to reading instruction, with students answering questions about the text and teacher-led class discussions focused on text content, although business-as-usual also contained elements prevailing in the aforementioned alternative instruction programs (e.g., activating prior knowledge, explaining vocabulary). Both types of control groups occurred with all types of outcome measures.

Stepwise model fitting showed that effect sizes were not systematically different when the intervention was compared to an alternative program instead of to business-as-usual; questions ($\Delta\chi^2(1) = 1.18, p = .28$), recall ($\Delta\chi^2(1) = 0.44, p = .51$), summarization ($\Delta\chi^2(1) = 2.16, p = .14$), and text structure knowledge ($\Delta\chi^2(1) = 1.14, p = .29$). For example, the estimated effect size of text structure instruction

on recall was $g = 0.32$ ($SE = 0.06$) when compared to an alternative instruction, and $g = 0.24$ ($SE = 0.12$) when compared to business-as-usual, which was not statistically significant. For the remaining analyses, we therefore did not distinguish between the two types of control conditions.

Standardized and non-standardized measures

Both standardized and non-standardized measures were used to evaluate the effects of text structure instruction. For comprehension questions, approximately 22% of the effect sizes were based on standardized measures. For recall, summarization, and text structure knowledge tests, only non-standardized measures were used. We could not demonstrate an effect of standardization ($\Delta g = -0.02$, $SE = 0.11$; $\Delta\chi^2(1) = 0.04$, $p = .84$). Differences in effect sizes due to text structure instruction were similar when measured with standardized or non-standardized comprehension questions.

Immediate and delayed effects

Not all studies provided data on the maintenance of effects. Delayed posttests were administered in approximately one third of the studies ($n = 16$). Most of these concerned comprehension questions ($n = 12$) and recall ($n = 6$) and took place in the second or third week after completing the intervention (63%). All immediate effect sizes were above 0.20, and therefore meaningful: students who received text structure instruction outperformed the control group at questions ($g = 0.25$, $SE = 0.07$), recall ($g = 0.38$, $SE = 0.06$), summarization ($g = 0.58$, $SE = 0.09$), and text structure knowledge ($g = 0.34$, $SE = 0.03$).

However, the delayed effects *decreased* significantly; questions ($\Delta\chi^2(1) = 6.89$, $p < .001$), recall ($\Delta\chi^2(1) = 4.83$, $p = .03$), summarization ($\Delta\chi^2(1) = 9.73$, $p = .002$). For each of the outcome measures, the differences between groups with or without text structure instruction completely disappeared at the delayed posttests (Table 3), as none of the estimated effect sizes reached significance; for questions ($g = -0.05$, $SE = 0.11$), for recall ($g = 0.13$, $SE = 0.11$), and for summarization ($g = -0.06$, $SE = 0.20$).

TABLE 3Immediate and delayed effect sizes (ESs) in Hedges' g (and SE)

Outcome measure	Immediate ES	Delayed ES
Comprehension questions	0.25 (0.07)	−0.05 (0.11)
Recall	0.38 (0.06)	0.13 (0.11)
Summarization	0.58 (0.09)	−0.06 (0.20)
Text structure knowledge	0.34 (0.03)	

Note. The delayed ESs show the difference between experimental and control groups at delayed posttests. Due to power insufficiencies ($n = 3$), no delayed effect sizes were calculated for text structure knowledge.

In other words, although the students in the text structure condition outperformed the controls on each outcome measure at the immediate posttest, this difference between conditions was not maintained at the delayed posttest, where students' performance in both conditions was similar. In most cases, the performance of the experimental groups showed a stronger decrease after the immediate posttest than the control groups, whose performance remained rather stable, or showed a small decrease.

The remaining effect sizes were still heterogeneous after adding the previously mentioned variables ($Q_{Questions}(87) = 300.18, p < .001$; $Q_{Recall}(83) = 245.08, p < .001$; $Q_{Summarization}(50) = 276.73, p < .001$; $Q_{Knowledge}(25) = 79.26, p < .001$). We decided to conduct moderator analyses with content-related and instructional variables.

Text variables

We examined whether the type and number of text structures affected the effect sizes, by adding genre and the number of structures taught to the four models.

Genre

Most studies discussed the effects of informative text structure instruction ($n = 34$). In some studies, students also received instruction in narrative story structure ($n = 5$) or in narrative story structure only ($n = 5$). Interventions on informative text structures included all types of outcome measures, whereas interventions

involving narrative structure mainly used comprehension questions and recall. We could not demonstrate an effect of genre on these outcome measures; questions ($\Delta\chi^2(1) = 0.12, p = .73$), and recall ($\Delta\chi^2(1) = 0.11, p = .74$). There was also no interaction effect of genre on delayed posttest performance; questions ($\Delta\chi^2(1) = 0.89, p = .34$), and recall ($\Delta\chi^2(1) = 0.11, p = .74$). Interventions focusing on informative text structures only versus studies (also) including narrative texts were comparable in terms of their effects on comprehension. As the number of studies focusing on narrative story structure was limited, we could not examine genre effects on summarization and text structure knowledge.

Number of text structures

Some interventions studies focused on only one text structure ($n = 20$), other studies taught up to five structures ($n = 9$). Overall, the description and compare-contrast structure were taught most frequently. More structures were taught in interventions with a text structure knowledge test ($M = 3.81, SD = 1.47$) or summarization task ($M = 2.37, SD = 1.44$) as outcome measure than in interventions with questions ($M = 1.53, SD = 1.07$) or recall tasks ($M = 1.64, SD = 1.35$). There was a small negative effect of the number of different text structures on text structure knowledge ($\Delta g = -0.06; \Delta\chi^2(1) = 4.64, p = .03$), but not on questions ($\Delta\chi^2(1) = 0.54, p = .46$), recall ($\Delta\chi^2(1) = 0.003, p = .96$) or summarization ($\Delta\chi^2(1) = 0.21, p = .65$). The more different text structures were taught during an intervention, the lower the scores on the text structure knowledge test (i.e., $\Delta g = -0.06$ times the number of different text structures taught). We could not show a curve linear effect. Number of text structures taught did not matter when students made a summary, answered comprehension questions or carried out a recall task.

We also checked for an interaction effect between the number of text structures taught and measurement moment to see if the number of text structures mattered for delayed posttests. For none of the outcome measures, this interaction effect was found; questions ($\Delta\chi^2(2) = 3.44, p = .18$), recall ($\Delta\chi^2(2) = 0.02, p = .99$), summarization ($\Delta\chi^2(2) = 0.21, p = .90$), or text structure knowledge ($\Delta\chi^2(1) = 4.57, p = .03$). The number of different text structures taught did not affect maintenance effects.

Content-related features and instructional components

We analyzed whether the effects of text structure instruction were affected by the content features and instructional components listed in Table 1. Because of a limited number of observations, we did not perform this moderator analysis for text structure knowledge.

All final models were an improvement relative to the models without moderating content-related and instructional variables; questions ($\Delta\chi^2(5) = 14.13$, $p = .015$), recall ($\Delta\chi^2(5) = 15.49$, $p = .008$), and summarization ($\Delta\chi^2(3) = 20.81$, $p = .001$). The four final models explain 25-27% of the variance in effect sizes.

Content-related features

Table 4 shows the estimated immediate effects per outcome measure, and the estimated additional effects of various content features and instructional components. The parameter estimates from the final models (see Appendix A) show that not all features contributed evenly to the effects on the different outcome measures. Training in text structure recognition had a significant effect on question answering ($g = 0.98$, $SE = 0.30$, $p = .001$) and recall ($g = 1.03$, $SE = 0.39$, $p = .009$). For recall, instruction on paragraph-level structure also mattered ($\Delta g = 0.57$, $SE = 0.29$, $p = .03$), but this was not the case for comprehension questions. For summarization skills, it was specifically a focus on paragraph-level structure that mattered ($\Delta g = 0.91$, $SE = 0.22$, $p < .001$), whereas training in only top-level text structure recognition did not significantly improve students' summarization skills ($\Delta g = 0.22$, $SE = 0.40$, $p = .58$). Apparently, when it comes to summarizing, students benefit most from text structure instruction that also focuses on the internal structure of paragraphs ($g = 0.91 + 0.22 = 1.13$).

General attention to structure visualizations had no demonstrable impact on students' performance on text comprehension questions, unless intervention programs emphasized the actual construction of graphic organizers and story maps ($\Delta g = 0.39$, $SE = 0.15$, $p = .009$). For recall, simple exposure to graphic organizers had a negative effect ($\Delta g = -0.44$, $SE = 0.18$, $p = .02$), whereas active construction had a significant positive effect ($\Delta g = .51$, $SE = 0.17$, $p = .002$). When students actively created or filled out maps or graphic organizers in addition to text structure recognition, this had an effect on comprehension questions ($g = 0.64$) and recall ($g = 1.03$).

Structure-based summarization training in general had no demonstrable additional effect on question answering, recall or summarization skills, over and above training in text structure recognition. However, specific training in the rule-based summarization technique positively affected summarization skills ($\Delta g = 0.64$, $SE = 0.21$, $p = .005$) and recall ($\Delta g = 0.34$, $SE = 0.12$, $p = .004$), but had no significant impact on comprehension questions. When interventions trained students to apply a fixed set of structure-based rules to summarize text, this resulted in net immediate effects of $g = 0.73$ on summarization and $g = 1.12$ on recall. None of the content-related features had a demonstrable impact on delayed posttests. As was reported earlier, the differences between students receiving text structure instruction and students in the control condition disappeared at the delayed posttests.

TABLE 4
Immediate effects (Δg) for content-related and instructional variables

	Outcome measure		
	Questions	Recall	Summarization
Text structure recognition (intercept)	0.98	1.03	0.22 (<i>ns</i>)
Paragraph-level structure recognition (Δg)		$\Delta 0.57$	$\Delta 0.91$
Structure visualization		$\Delta -0.44$	
Active structure mapping (Δg)	$\Delta 0.39$	$\Delta 0.51$	
Structure-based summarization			
Rule-based summarization (Δg)		$\Delta 0.12$	$\Delta 0.21$
Teacher: No modeling			
Student: No individual practice			

Note. *ns* = not statistically significant. Empty cells are nonsignificant.

Instructional components

There was no demonstrable additional effect of instructional features on immediate measures. Interventions including teacher modeling or individual student practice resulted in similar effect sizes as interventions with only explicit instruction or collaborative activities. However, instructional components may have an effect on delayed effect measures. There was a marginally significant effect of individual practice on the delayed effect on questions ($\Delta\chi^2(4) = 9.39, p = .052$)¹.

That is, students performed worse on comprehension questions during delayed posttest ($\Delta g = -1.04, SE = 0.39, p = .007$) when interventions lacked individual activities. With individual practice during the intervention program, the delayed effect for text structure instruction on questions would be $g = 0.82$, instead of $g = -0.23$ without individual practice. This finding suggests that individual activities might work as protecting factor against the relapse in scores on delayed posttests.

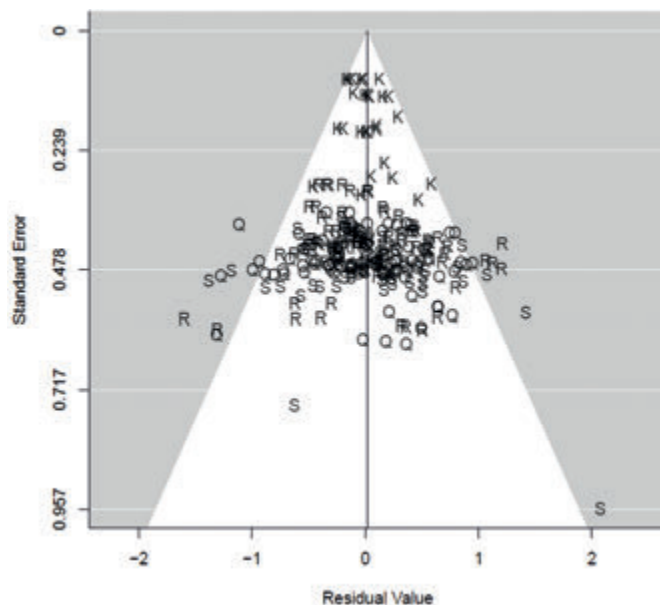
Funnel plot

Figure 1 shows the funnel plot of the four final models combined. The residuals of the final models with the explanatory variables are plotted against the standard errors. Most points were located in the region between the straight lines. For all outcome measures taken together, only 16 effect sizes (5.7% of the total sample) were identified as outliers. In both the lower bound and upper bound outliers, effect sizes were based on various outcome measures, although question effect sizes were slightly overrepresented as lower bound outliers. It shows that there was still a significant amount of unexplained heterogeneity in the four data sets $QE_{Questions}(82) = 272.78, p < .001$; $QE_{Recall}(76) = 219.37, p < .001$; $QE_{Summarization}(45) = 193.31, p < .001$; $QE_{Knowledge}(22) = 59.47, p < .001$, but there did not seem to be any systematic heterogeneity.

¹ This effect of individual practice is worth mentioning, even though it is not significant. A more powerful analysis in which the delayed effects on questions, recall, and summarization were collapsed and task type was modeled as an interaction effect, at this point showed a significant improvement in model fit ($\Delta\chi^2(2) = 6.84, p = .03$), as well as a significant parameter estimate for individual practice on delayed effect measures ($\Delta g = -0.74, SE = 0.31, p = .02$).

FIGURE 1

Funnel plot. Q: questions, R: recall, S: summarization, K: text structure knowledge



2.5 Conclusion and Discussion

Immediate and delayed effects of text structure instruction

Our study reveals that text structure instruction can improve reading comprehension skills in grades 4-6. Narrative and expository interventions seem equally effective in this age group. Contrary to previous meta-analyses (Hebert et al., 2016; Pyle et al., 2017), we could not demonstrate a difference between standardized and non-standardized tests, but our meta-analysis does show that the type of outcome measure (comprehension questions, recall, summary, text structure knowledge) has a dramatic impact on effect sizes. When compared to regular reading comprehension instruction, text structure instruction has an overall immediate effect of $g = 0.25$ on comprehension questions, $g = 0.38$ on recall, $g = 0.58$ on summarization, and $g = 0.34$ on text structure knowledge. However, at delayed posttests, the differences between groups that received text structure instruction or regular reading instruction could no longer be demonstrated.

At first glance, the overall effect of text structure instruction on comprehension questions might seem relatively low ($g = 0.25$), but this constitutes

an additional effect over and above business-as-usual gains in reading comprehension. Moreover, the effect sizes need to be evaluated in context: they were obtained in authentic educational contexts and not in controlled lab settings, so that effect sizes of around .20 that are often classified as small, are actually of policy interest (Durlak, 2009; Hedges & Hedberg, 2007). Furthermore, the effect size on comprehension questions is similar to the effect of other educational interventions (e.g., $g = 0.24$ in Lipsey et al., 2012). Therefore, text structure instruction should be considered as a way to support the reading-to-learn transition, so that the persistent 'fourth-grade slump' effect can be reduced (Chall & Jacobs, 1983).

We found that text structure instruction had a similar additional effect when compared to either business-as-usual reading instruction or a heterogeneous subset of alternative interventions (e.g., vocabulary instruction, cognitive reading strategies). This does not imply that in general these alternative interventions are not better than business-as-usual instruction. In fact, various meta-analyses have demonstrated the positive effect of vocabulary instruction (Stahl & Fairbanks, 1986) and (meta)cognitive learning strategies (e.g., Donker et al., 2014; Okkinga et al., 2018a) on comprehension. Our result might be due to the fact that the control condition consisted of a very heterogeneous subset of alternative interventions, resulting in a baseline reference group that is quite comparable to business-as-usual instruction.

The effect size on comprehension questions ($g = 0.25$) was lower than on the other outcome measures; on tests in which students had to apply their knowledge of text structure more directly (e.g., recognizing text structures), the immediate effect size was $g = 0.34$. Although comprehension questions are presented as one outcome measure, they can target various aspects of a text, and thereby might still measure different things (Keenan et al., 2008; Nation & Snowling, 2011). As text structure instruction provides students with the knowledge and tools to process text into more coherent and organized mental schemata, it seems to matter whether comprehension questions involve surface code, text-base, or situation model comprehension skills (Kintsch, 1988; 2004; 2013). Unfortunately, most studies did not specify whether comprehension questions concerned situation-model questions or local text issues that might be

answered without full understanding at the level of the situation model (van den Broek & Kremer, 2000).

Tasks that tap more into situation model comprehension - such as recall and summarization - might be more suitable candidates for evaluating effects of text structure instruction. Indeed, summary and recall measures yielded larger overall effects of text structure instruction ($g = 0.37$ and $g = 0.57$, respectively), which is in line with the meta-analysis by Pyle et al. (2017) who showed that the effect of text structure instruction was larger on graphic organizer tasks than on questions. However, even for summarization and recall tasks, it is important to note that they yield qualitatively different results: summarization tasks typically evoke more main ideas, whereas recall tasks might evoke details as well (Riley & Lee, 1996).

Our meta-analysis shows that outcome measure matters for evaluating the effectiveness of text structure interventions. As each type of outcome measures might rely on a slightly different constellation of comprehension skills and thereby measure different aspects of the reading process (Brown et al., 1983; Hare & Borchardt, 1984), future text structure research should include multiple outcome measures (Bohaty et al., 2015), as in the studies by Meyer et al. (2002, 2011) and Wijekumar et al. (2012, 2013, 2014). Another important step would be to disentangle how the different outcome measures for reading tap into the numerous skills involved in comprehension processes (Graesser et al., 1994; Keenan et al., 2008), so that reading interventions can be evaluated more adequately.

The students in the experimental condition no longer outperform the control group at delayed posttests. This resonates with the findings by Hebert and colleagues (2016) that delayed effects of text structure instruction are much smaller and less consistent than immediate effects. In fact, we found that in many studies the performance of the experimental group decreased between immediate and delayed posttest, whereas the control group performed rather similar on immediate and delayed posttests, or showed a small decrease as well.

A methodological factor that might contribute to this finding, is the fact that delayed posttests sometimes required transfer when students were tested on untaught text structures. Due to a limited number of studies with delayed posttests, we could not examine this transfer effect, but Hebert and colleagues (2016) have shown much smaller effects in far-transfer cases. Another explanation

for the lack of maintenance is the fact the intervention studies made no effort to promote maintenance; the highlights of text structure instruction were not repeated in the period between immediate and delayed posttest. Finally, the quality of business-as-usual instruction might be insufficient to help students to maintain their newly acquired knowledge about text structures. Teachers in the upper elementary grades often fail to employ evidence-based approaches (Duke et al., 2011; Wijekumar et al., 2019a), and struggle themselves with text structure recognition (Beerwinkle et al., 2018; *Chapter 4* – Bogaerds-Hazenberg et al., 2019; Reutzel et al., 2016) and main idea identification (e.g., Kucan et al., 2011). Also, curricular materials for reading in grades 4 and 5 hardly include text structure instruction (Beerwinkle et al., 2018; Bogaerds-Hazenberg et al., 2017).

Content-related and instructional components moderating the effects

The ability to recognize text structures is beneficial for increasing reading comprehension, irrespective of outcome measure and genre. Our meta-analysis also shows that the effects of moderating content-related and instructional variables are different per outcome measure, and moves beyond the question *what* elements to include in text structure instruction (e.g., graphic organizers) by providing some insight into *how* to include and refine these elements (e.g., how to offer graphic organizers).

Content-related features

When it comes to effect sizes for summarization and recall, students particularly benefit from text structure instructions that focus on the paragraph-level ($g = 1.60$ on recall and $g = 1.13$ on summarization), and not just on top-level structures. This corroborates the claim that successful summarization training combines text structure recognition and main idea identification within paragraphs (Miyatsu et al., 2018). The large effect of text structure instruction on summarization supports the hypothesis that text structure provides students with the necessary knowledge and tools to distinguish important from unimportant information (Hogan et al., 2011; Meyer et al., 1980; Taylor, 1985; Winograd, 1984), and that it helps them to see how these main ideas are organized at a higher level (Miyatsu et al., 2018; Westby et al., 2010). Furthermore, it can function as a mnemonic aid that improves text recall (Taylor, 1982). Alternatively, it might be easier to establish differences

in main ideas at the paragraph-level than at the text-level, as a paragraph-level focus invites students to produce multiple main ideas, instead of a single main idea for the text as a whole. As all interventions included instruction about signaling words, we could not analyze whether a focus on signaling words moderated the effects of text structure instruction.

Structure-based summarization also improves students' text comprehension, especially when students learn specific rules and tricks for paraphrasing the main idea (e.g. rule-based summarization). This yielded net immediate effects of $g = 1.15$ for recall and $g = 0.43$ for summarization. This corroborates the idea that explicit knowledge about text structure (e.g., structure-specific questions, signaling words) can provide students with useful tools to identify and formulate main ideas, and to reorganize these in a coherent way (Elledge, 2013; Meyer et al., 1980; Miyatsu et al., 2018; Stevens, 2018; Taylor, 1985; Westby et al., 2010), whilst eliminating redundant information (Brown et al., 1983; McNeil & Donant, 1982). While structure-based summarization techniques were not included as moderating variable in previous meta-analyses (Hebert et al., 2016; Pyle et al., 2017), we show that it is an important ingredient of successful text structure interventions. Summarization strategies based on the internal structure of paragraphs seem more helpful in improving students' performance than strategies based on external markers of text structure (i.e., headings and subheadings), possibly because the former provide students with the necessary skills to distill main ideas from the text, even in the absence of unambiguous external markers of text structure.

Structure visualizations are often part of larger text structure strategy interventions, but are also used in various studies on graphic organizers and story mapping. Over the years, graphic organizer research has generated mixed results, which has often been attributed to the types of graphic organizers used, to the level of instructional support that was provided, or whether they were used as pre-reading or post-reading activity (Griffin et al., 1995; Jiang & Grabe, 2007). Although previous meta-analyses suggest that graphic organizers might increase the effectiveness of text structure interventions (Hebert et al., 2016; Pyle et al., 2017), they did not examine their presence as a moderating variable. Our meta-analysis shows that the inclusion of structure-based visualizations has positive effects on comprehension and recall, as long as students actively fill out these

maps and graphic organizers. Simple exposure has no demonstrable effects, or even a negative effect on recall. It seems crucial that text structure instruction provides ample opportunities for students to practice filling out structure-based graphic organizers and maps after teacher-led instruction. This underscores the importance of an instructional approach that displays a gradual release of responsibility from teacher to student (e.g., Fisher & Frey, 2021).

The importance of the active construction of graphic organizers contradicts the conclusion of Stull and Mayer (2007), who found that constructing graphic organizers increased the extraneous cognitive processing load and interfered with learning. In the studies that were included in our meta-analysis, we found that 'constructing graphic organizers' consisted of students filling out missing information in an already given graphic organizer, and did not require them to draw the whole structure by themselves (i.e. choosing the right boxes, arrows). Possibly, this is a less complex task for students, as the text structure is scaffolded in the graphic organizer, and students' only concern is to find the right ideas to put in the boxes. This type of task might reduce extraneous load, but still fits with the theory that deep learning occurs when students are encouraged to engage in *productive* learning activities (Mayer, 2003, 2004). So even finishing a partially completed organizer already provides an opportunity for deep text processing (Jiang & Grabe, 2007). With current technological trends and development of mapping software, it seems relevant to explore digital opportunities for incorporating more learning activities focused on text structure visualizations.

Although previous meta-analyses have shown that teaching multiple text structures has a positive impact on students' performance (Hebert et al., 2016; Pyle et al., 2017), we could not demonstrate this effect. Only for text structure knowledge we found that students' performance tended to be slightly lower when they encountered more different structures. This seems logical, as students need to remember more different types of structures on these text structure knowledge tests. Still, teaching multiple structures has at least no demonstrable *negative* impact on the other comprehension measures. Therefore, we believe that teaching multiple text structures is useful, as students can learn from comparing and contrasting the characteristics of various structures, and become more aware of the differences between structures, and possibly even transfer knowledge to untaught text structures (Hebert et al., 2016). Also from a practical perspective, it

is important for students to recognize more than one structure, as most (educational) texts are a combination of multiple text structures nested within one another (Jiang & Grabe, 2007). As Pyle et al. (2017) have also suggested, we believe that it is worthwhile to engage in further research that addresses the order and complexity of different text structures that are taught.

Instructional components

Although it matters *how* a skill is taught to students, most meta-analyses on literacy research do not evaluate the effect of instructional components such as modeling or collaborative practice, possibly because intervention descriptions often extensively focus on content. As Pyle et al. (2017) have pointed out, the term *explicit instruction* is often used in describing text structure interventions, but is in itself very broad in terms of instructional features present (Archer & Hughes, 2011).

Despite the limited descriptions of the instructional approach in most research articles, our study suggests that the instructional approach might moderate the effects: interventions with individual student activities generated slightly higher effects on comprehension questions at delayed posttests. Although several instructional models hypothesize practice with peers to be an important step in the gradual release of responsibility from teacher to student (Fisher & Frey, 2021; Pearson & Gallagher, 1983), we could not demonstrate a moderating effect of collaborative activities. Due to the poor description of instructional features in most studies, it is hard to interpret this finding. Students who do not practice alone but only with their peers may not fully acquire the skill they have to learn, easily forget it, and therefore fail on delayed posttests. Alternatively, activities that were labeled as ‘collaborative activities’ in the intervention might not actually have met the criteria for effective cooperation (see Johnson & Johnson, 1984; 2017).

Several researchers have expressed the need to pay more attention to the fidelity of implementation in text structure intervention studies (Bohaty et al., 2015). We believe that the quality of instructional components should also be included in such evaluations. More specifically, parallel to the recommendations made by writing researchers (Bouwer & De Smedt, 2018), future reading research articles should systematically provide details on the intervention context and on the design principles of the intervention, at both a macro-level (i.e., focus and mode of instruction, sequencing of content) and micro-level (e.g., instructional

activities, learning activities, materials). This will increase the transparency of intervention results and might promote the implementation of concrete activities in educational contexts (Fidalgo et al., 2018). The *Gradual Release of Responsibility* model (Fisher & Frey, 2021; Pearson & Gallagher, 1983) provides a useful framework to more systematically describe, test and evaluate the quality of instructional components.

Given that we know from various meta-analyses that text structure instruction is effective (the 'what'), for students of various ages (the 'when'), it is important that future studies focus on instructional practice (the 'how'). Now is the time to examine the effectiveness of a greater variety of instructional features in the context of reading instruction, so that we can ameliorate the context in which text structure instruction is given (Williams, 2018). A first attempt has already been made by Meyer et al. (2002, 2010) and Wijekumar et al. (2012, 2013, 2014), who try to unravel the effects of providing for instance individualized feedback or the effects of tutoring in the context of a web-based intervention. Recently, qualitative research has been undertaken to qualitatively describe teacher instructional practices and pedagogical content knowledge in the context of text structure interventions (Beerwinkle et al., 2018; *Chapter 4*; Wijekumar et al., 2019a), which provides more insight into the instructional components that influence intervention success.

In sum, this meta-analysis shows that text structure instruction has a positive effect on students' reading comprehension skills over and above regular reading programs: it improves their performance on comprehension questions, recall, summarization tasks and has a positive effect on their text structure knowledge. However, at delayed posttests differences between experimental groups and control groups can no longer be demonstrated. Hence, it seems a promising avenue to incorporate text structure instruction into primary school curricula so that students' comprehension skills can be strengthened and positive effects can be maintained.

CHAPTER 3



What textbooks offer and what teachers teach: An analysis of the Dutch reading comprehension curriculum

Abstract

This chapter presents the current practice of reading comprehension instruction in the Netherlands. Although science has made significant progress in unraveling effective practices for reading comprehension instruction, questions have been raised about the quality of the current Dutch reading comprehension curriculum. In a mixed-method study, we analyzed the content and pedagogy in textbooks for reading comprehension instruction (i.e., the implemented curriculum) and teachers' evaluation and use of these books (i.e., the enacted curriculum). The data are based on a materials analysis of reading comprehension lessons ($n = 80$) in eight textbooks for grades 4 and 5. This analysis was complemented with semi-structured teacher interviews ($n = 29$) and lesson observations ($n = 11$), with a focus on the quality of reading strategy and text structure instruction in the curriculum. Main findings are (1) a lack of alignment between lesson goals, theory, and assignments, (2) a strong focus on practicing strategies, (3) limited declarative knowledge about strategies and text structure, (4) little opportunities for self-regulated strategy application, and (5) strong emphasis on individual question answering. This chapter concludes with several recommendations for improving the current curriculum.

This chapter is an extended version of:



Bogaerds-Hazenbergh, S.T.M., Evers-Vermeul, J., & van den Bergh, H.H. (2022). What textbooks offer and what teachers teach: an analysis of the Dutch reading comprehension curriculum. *Reading and Writing*, 35(7), 1497-1523. <https://doi.org/10.1007/s11145-021-10244-4>

3.1 Introduction

Reading comprehension skills are crucial for children's school success (e.g., Chall & Jacobs, 1983), and their future social lives and careers (Oakhill et al., 2015; Snow, 2002). Reading comprehension is however a multifaceted and complex skill that is affected by many underlying components. For instance, various studies have shown that text comprehension is affected by readers' prior knowledge (Kendeou & van den Broek, 2007; Smith et al., 2021; van Moort et al., 2020), vocabulary breadth and depth (Beck et al., 1982; Stahl & Fairbanks, 1986), metacognitive skills (Soodla et al., 2017; Soto et al., 2019), knowledge of reading strategies (Bimmel, 2001; Cain, 1999; Droop et al., 2015; Houtveen & van de Grift, 2007), and knowledge of text structures (*Chapter 2* – Bogaerds-Hazenberg et al., 2021; Hebert et al., 2016; Pyle et al., 2017).

The importance of reading comprehension skills calls for high-quality reading instruction. Textbook quality is known to play a crucial role here (Aaron et al., 2008; Wijekumar et al., 2021). In the Netherlands, teachers use specific reading comprehension textbooks, as reading comprehension is typically taught as a separate subject – apart from language teaching and technical reading (Garbe et al., 2016; van Gelderen & van Schooten, 2011). Where some countries (e.g., Austria, China, Portugal) prescribe the core contents of the reading curriculum and sometimes even require schools to use textbooks that are explicitly approved by their government (Mullis et al., 2016; Zhang et al., 2021), the Dutch government only prescribes attainment targets in the so-called core objectives of the Reference Framework for language and arithmetic (Committee Meijerink, 2008). Therefore, educational publishers and schools can freely determine which reading strategies are taught and when (Garbe et al., 2016), and what instructional approach is used (Bruggink & Netten, 2017).

The quality of the current Dutch reading comprehension curriculum is under debate. Both national and international assessments show that a considerable group of Dutch students struggle with reading comprehension, both during primary education (e.g., PIRLS-2016: Gubbels et al., 2017; PPON-54: Kuhlemeier et al., 2014) and beyond (PISA-2015: Feskens et al., 2016; PISA-2018: Gubbels et al., 2019). The problem seems persistent and has not really changed over the past few years. In fact, a growing number of students are at risk of ending up as low literate (Pereira & Nicolaas, 2019). Also, Dutch students display the lowest levels of

reading motivation in the world, and feel less involved in reading instruction than students in other OECD countries (Gubbels et al., 2017; Meelissen et al., 2012). Some institutions have therefore expressed their doubts about the effectiveness and quality of current teaching materials, and recommend working with evidence-based pedagogies and materials (Dutch Inspectorate of Education, 2019; Pereira & Nicolaas, 2019).

All in all, a gap arises between the intended curriculum (i.e., educational objectives) and the attained curriculum (i.e., students' achievements) (Scheerens, 2017; Valverde et al., 2002). To some extent, this might be related to the quality of teaching materials. Textbooks form a crucial stepping stone between the intended and attained curriculum, and play a major role in the chain of effective instruction (Penuel et al., 2014). They affect teachers' curriculum enactment and classroom practice (Ball & Cohen, 1996), also with respect to reading instruction (Aaron et al. 2008; Dewitz & Jones, 2013; Harwood, 2017).

According to the Component Model of Reading Comprehension (Aaron et al., 2008), ecological components, such as the quality of textbooks and teacher knowledge, affect ultimate reading achievement. Teaching materials constitute the implemented curriculum, which entails the learning content and instructional approach that teachers offer in their classrooms. Their quality and alignment with government objectives strongly determine effective education (Dockx et al., 2020; Penuel et al., 2014; Scheerens, 2017).

Given the importance of high quality textbook content for reading comprehension, a recent research initiative examined textbook content in various countries (Wijekumar et al., 2021). Key findings were the limited attention for expository texts and text structure knowledge (e.g., Austria: Seifert, 2021; China: Zhang et al., 2021; Croatia: Peti-Stantić et al., 2021) and the strong focus on the recall of textual content in both instruction and activities, often at the expense of promoting higher-order thinking strategies such as inferencing (e.g., Malta: Aguis & Zammit, 2021; Portugal: Cordeiro et al., 2021; US: Beerwinkle et al., 2021). The overall conclusion of this cross-country textbook analysis was that most textbooks do not sufficiently incorporate evidence-based practices (Wijekumar et al., 2021).

Although the aforementioned textbook studies provide valuable insights into textbook quality, teachers' self-reported and actual use of these textbooks was not included, even though it is known that teachers moderate the effect of the

teaching materials they use (Brown, 2009; Harwood, 2017). They do so by selecting the materials they teach, but mostly by the level of pedagogical content knowledge they can build upon in their enactment of these materials (PCK; Blömeke et al., 2014; Gudmundsdottir & Shulman, 1987; Scheerens & Blömeke, 2016). Even an excellent quality of teaching materials is therefore no guarantee that teachers are able to use them. This has also been shown in the area of reading comprehension instruction (Dewitz & Jones, 2013; Duke et al., 2011; Valencia et al., 2006). Therefore, textbook research needs to be complemented with an analysis of the enacted curriculum which refers to teachers' actual use of these materials (Brown, 2009; Harwood, 2017), which is strongly affected by textbook quality (Valencia et al., 2006), and by teachers' knowledge, habits, and beliefs (e.g., Ball & Cohen, 1996; Brown, 2009; Penuel et al., 2014).

Despite the debate on the quality of the Dutch reading curriculum (Dutch Inspectorate of Education, 2019; Pereira & Nicolaas, 2019), no recent studies have been conducted that closely examine both the implemented and enacted curriculum (Hoogeveen, 2018). There are indications that, in most schools, teachers still dutifully enact their textbooks, often without explicitly reflecting on learning goals and curricular structure (Scheltinga et al., 2013). This puts them at risk to copy flaws if textbooks lack quality (Aarnoutse, 1990). For example, some studies suggest that too much emphasis is put on question answering, at the expense of improving students' reading process (Bonset & Hoogeveen, 2009; Rooijackers et al., 2020; 2021), that both teachers and students often seem to consider reading comprehension as 'answering questions about texts' (Berends, 2011), and that teachers often foreground the topic of the text, instead of the reading skills to be acquired (Wieberdink & Kuster, 2015).

The aim of the current mixed-methods study is to carefully describe the content and instructional features of teaching materials that are used for reading comprehension instruction in Dutch primary education (the implemented curriculum), and examine how teachers experience and use these in their classrooms (the enacted curriculum), with a specific focus on reading strategy and text structure instruction.

3.2 Theoretical Framework

Although the development of reading comprehension depends on a complex constellation of many different underlying skills, our current study puts emphasis on two aspects: text structure instruction and the teaching of reading strategies. Both text structure and strategy knowledge play a crucial role in theoretical models on reading comprehension such as the Landscape Model (van den Broek et al., 1999) and the Construction-Integration Model (Kintsch, 1988, 2004, 2013; van Dijk & Kintsch, 1983), and have been identified as powerful ingredients of comprehension instruction (*Chapter 2*; Duke et al., 2011; Pyle et al., 2017). Text structures and reading strategies should therefore be well represented in curricular materials (Wijekumar et al., 2021).

First, the Landscape Model theorizes that readers have to distribute their limited attentional resources over different concepts in a text to keep them activated, and to connect them to their prior knowledge. While building mental representations of the text out of these network activation patterns, readers need to monitor whether they are realizing their reading goals, or if they have a comprehension problem that needs to be resolved first. Metacognitive skills and sufficient knowledge about reading strategies are crucial to cope with comprehension problems (van den Broek et al., 1999). Therefore, reading strategies – mental tools that readers purposively use to monitor, repair, or support comprehension – often form a crucial part of reading curricula (Aarnoutse, 2017).

Teacher-directed reading strategy instruction can positively affect students' strategy use and knowledge, and improve reading comprehension performance (Bråten & Anmarkrud, 2013; Brown et al., 1996; Houtveen & van de Grift, 2007; Okkinga et al., 2018a; Palincsar and Brown, 1984; Soodla et al., 2017). It therefore seems important to teach primary school students how to monitor and resolve comprehension problems by using a repertoire of evidence-based strategies, such as predicting information, questioning, summarizing, and inspecting textual features (Aarnoutse, 2017; Bimmel, 2001; Block & Duffy, 2008; Duke et al., 2011).

However, intervention studies do not always show substantial effects of strategy instruction on standardized tests (e.g., Aarnoutse & Schellings, 2003; Andreassen & Bråten, 2011; Droop et al., 2016; Muijselaar et al., 2018; Okkinga et al., 2018b). Besides methodological explanations (short interventions, test-related

issues), it also seems that the quality of reading strategy instruction can break or make an intervention (Aarnoutse & Schellings, 2003; Andreassen & Bråten, 2011). For example, Okkinga et al. (2018b) showed that the more elaborate explanations teachers provided on the nature, function, importance, and application of reading strategies, the better students performed on reading comprehension tests. Similarly, Droop et al. (2016) showed that investing in declarative knowledge about strategies without instruction in how and when to use these strategies does not directly pay off in terms of reading achievements for third and fourth graders.

In other words, reading strategies should be taught in a way that helps students to become strategic readers who engage in a goal-directed, deliberate, and self-regulated use of these strategies (Afflerbach et al., 2008; Alexander, 2018; Nash-Ditzel, 2014). One important suggestion is therefore that high-quality reading strategy instruction should enable students to develop three types of knowledge (Cross & Paris, 1988; Kostons et al., 2014; Paris et al., 1983):

- (1) Declarative knowledge: factual knowledge about strategies, text structures, and genres, needed for adequate goal setting and strategy use;
- (2) Procedural knowledge: knowing and reflecting on how to apply reading strategies while reading;
- (3) Conditional knowledge: when-and-why knowledge needed to match reading behavior to the task or text at hand; related to metacognition.

The more knowledge students obtain about how different reading strategies operate in various situations (e.g., reading purposes, text structures), the better they will adapt their reading behavior to a given situation, both inside and outside the direct context of comprehension lessons (Afflerbach et al., 2008; Alexander, 2018). For example, a reading strategy like skimming is useful to read a list of online search results, but not for reading poetry. Without sufficient conditional knowledge, students are unable to match their use of reading strategies to the text at hand, lack insight in the value of reading strategies, and cannot become self-regulated strategic readers (Nash-Ditzel, 2014; Simpson & Nist, 2000).

Second, teaching materials should include instruction about text structures, or the 'organization of ideas, the relationship among the ideas, and the vocabulary used to convey meaning to the reader' (Pyle et al., 2017:1). According

to the Construction-Integration Model of text comprehension (Kintsch, 1988; 2004; 2013), readers parse text input into concepts and relationships which they need to organize in associative networks. They need to make connections between ideas within the text, and with prior knowledge, until they arrive at a coherent mental representation of the text: the situation model (Kintsch, 1988). Text structure can facilitate the construction of a situation model, as good readers do not simply use a default list-strategy to organize information, but instead pick up text-structural cues and use these to organize main ideas accordingly (Meyer et al., 1980).

It has been shown that knowledge about text structures helps students to locate, recall and interconnect certain pieces of information (Meyer & Ray, 2011). Explicit text structure instruction relates to better comprehension, recall, and summarization skills (*Chapter 2*; Hebert et al., 2016; Pyle et al., 2017). Therefore, a high-quality reading curriculum should also include explicit text structure instruction that enables students to acquire declarative knowledge about different text structures (e.g., on the rhetorical goal of each structure and on its defining signaling words), as well as procedural knowledge (e.g., how to recognize different text structures). Knowledge about text structure also matters for conditional knowledge, because an increased awareness of different informational patterns of text can facilitate a purposeful and context-sensitive application of reading strategies. For example, the main idea of compare-contrast texts is phrased differently than that of a cause-effect text (Hoch & McNally, 2020; Oakhill et al., 2015; Stevens & Vaughn, 2021).

In sum, with regard to curricular content, teaching materials should offer a broad spectrum of declarative, procedural, and conditional knowledge about, and room for practice with reading strategies, and provide explicit text structure instruction. Furthermore, they should provide students with ample opportunities to deliberately plan, monitor, and evaluate their strategy use in increasingly complex texts, until these overt strategies eventually transform into covert skills as readers use them more or less automatically (Afflerbach et al., 2008). In this study, we therefore focus on the question to what extent this declarative, procedural, and conditional knowledge is realized in Dutch teaching materials for reading comprehension (the implemented curriculum), and how teachers enact and evaluate this (the enacted curriculum).

Following a similar line of reasoning, we examine certain aspects of the instructional approach. The Dutch PIRLS-2016 assessment showed that many students (44%) always or very often work independently on reading assignments. This might be problematic, as recent eye-tracking research shows that Dutch secondary school students heavily rely on these questions for text comprehension, and that they hardly apply any reading strategies spontaneously, unless they are prompted by questions (Rooijackers et al., 2020). Also, various evidence-based instructional approaches are not always implemented as intended in classrooms (e.g., Van den Branden et al., 2019), possibly due to limited teacher knowledge on powerful instructional approaches (Okkinga et al., 2018b; 2021), or because of a long tradition in which answering comprehension questions has become the central activity during lessons, in primary education (Aarnoutse, 1991; Berends, 2011; Bonset & Hoogeveen, 2009) and beyond (Rooijackers et al., 2020; 2021).

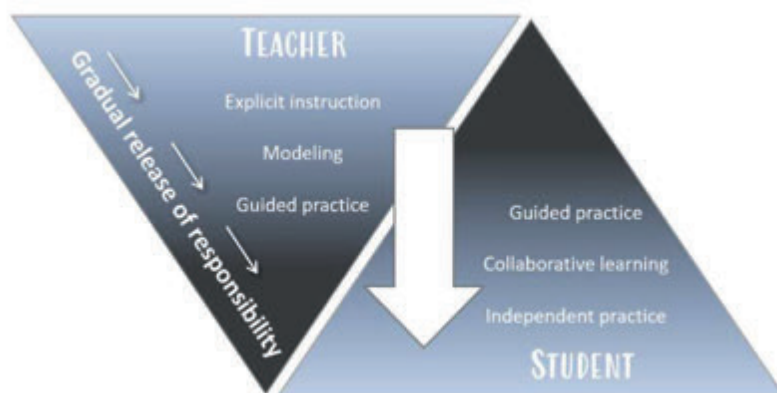
This raises the question what then constitutes an effective, evidence-based instructional approach to teaching reading comprehension. One influential model that comprises findings of various empirical studies on reading instruction is the Gradual Release of Responsibility model (GRR model; Fisher & Frey, 2021; Kelly, 2019; Pearson & Gallagher, 1983). This model describes the instructional practices for reading comprehension and their sequencing. The GRR model sequences instructional activities such that the responsibility for task completion relies first mainly on the teacher (i.e., through explicit instruction and modeling), and is then gradually transferred to students who work together until they reach independence (see Figure 1 for a visualization).

There is empirical back-up for many components of the GRR model. For example, positive effects have been found for explicit instruction combined with teacher modeling, when it comes to teaching reading strategies (Aarnoutse, 2017; Collins & Pressley, 2002) as well as text structures (Pyle et al., 2017). The process of modeling is a crucial step in comprehension instruction, as it makes invisible thinking processes visible to students (Duffy et al., 1986; Duke & Pearson, 2002; Schutz & Rainey, 2019). Since students have to worry less about text content while observing, working memory load is reduced and students can focus maximally on strategy application (Keehnen et al., 2015). In order to generate optimal modeling effects, teachers should include clearly marked moments of abstraction in which they highlight necessary components of strategy use or reflect on the reading

approach. This reflection or abstraction serves to disentangle the reading process from the immediate context of modeling (Rijlaarsdam, 2005; Schutz & Rainey, 2020).

FIGURE 1

Gradual Release of Responsibility instruction model: the triangles show the proportion of responsibility divided between teacher and students



After teacher-led instruction, the GRR model suggests student activities under supervision of their teacher (i.e., guided practice) and small-group collaboration (Duffy et al., 1986; Singer & Donlan, 1980), which also provides opportunities for reciprocal teaching and peer modeling (Palinscar & Brown, 1984), until students can work independently. Empirical studies have shown that high-quality small-group collaboration yields positive effects on comprehension (Palinscar & Brown, 1984; Puzio & Colby, 2013; Rosenshine & Meister, 1994; Vaughn et al., 2001), but also that individual practice is an important instructional component (*Chapter 2*).

However, there are indications that teachers find it difficult to implement the GRR model: they often struggle to provide small-group collaborative learning activities (Okkinga et al., 2018b; 2021; Spörer et al., 2009; Veldman et al., 2020), and often too quickly expect their students to implement reading strategies on their own (Almasi & Fullerton, 2012; Shanahan, et al., 2019). Therefore, in this study we also describe to what extent the GRR model is applied in Dutch teaching materials (implemented curriculum), and how teachers use and evaluate this blueprint for the instructional approach in practice (enacted curriculum).

3.3 Methods

In order to describe the Dutch reading comprehension curriculum (i.e., textbook quality, and teachers' perspectives and enactment of textbooks), we triangulated data from three different methods, thereby increasing the validity of our study (Table 1; Brown, 2009; Kuorikoski & Marchionni, 2016; Miles et al., 2013): a materials analysis, teacher interviews, and lesson observations. Appendix B provides additional information on the aspects that were coded across all three data sources.

Materials analysis

We made an inventory of the content of eight textbooks specifically designed for reading comprehension instruction (Table 1). Together, these books cover the vast majority of the Dutch curriculum and exemplify the diversity of the curricular landscape, as we included textbooks with very large market shares (e.g., *Nieuwsbegrip*) as well as two niche methods (*Bliksem*, *Lezen is weten*). All textbooks cover similar topics: students learn and practice comprehension strategies related to different phases of reading.

Per textbook, ten instruction lessons for grades 4 and 5 were randomly selected ($n = 80$; 12-18% of the year curriculum). The textbook analysis was complemented with an analysis of student worksheets, teacher manuals that typically contained the full blueprint for each lesson (e.g., modeling scripts, instructional approach), and informational brochures highlighting the intentions of the publishers. All materials were analyzed with a coding scheme that focused on content features of the lesson goals, theory and assignments, as well as on instructional features. This was a recursive process in which some categories were added to the initial coding scheme (Miles et al., 2013).

Coding of content features

We analyzed per lesson whether goals, theory and assignments related to declarative, procedural, and/or conditional knowledge. Most aspects were coded as dichotomous variables (present/not present in lesson); only assignments and questions were counted in order to examine the amount of practice.

TABLE 1

Overview of sources for the mixed-methods analysis per textbook

Textbook (<i>abbreviation and publisher</i>)	Materials analysis			Inter-views	Lesson obser-vation s
	Info bro-chures	Teacher manual	Text books, work sheets		
<i>Bliksem (BL)</i> (Bazalt, HCO, Expertisecentrum NL, 2013)	√	√	√	4	0
<i>Grip op lezen (GL)</i> (Malmberg, 2012)	√	√	√	4	2
<i>Kidsweek in de klas (KW)</i> (Young & Connected, 2009)		√	√	3	0
<i>Leeslink (LL)</i> (Malmberg, V-2020)	√	√	√	3	1
<i>Lezen in beeld (LB)</i> (Zwijssen, 2009)		√	√	4	0
<i>Lezen = weten (LW)</i> (Bazalt, 2011)	√	√	√	3	0
<i>Nieuwsbegrip XL (NB)</i> (CED-groep, V-2017)	√	√	√	4	6
<i>Tekst verwerken (TV)</i> (Noordhoff, 2006)	√	√	√	4	2

For the analysis of declarative knowledge, we coded the presence of factual knowledge about strategies (e.g., a list of strategies), explanation about the features of global text structures (e.g., cause-effect) and genres (e.g., a recipe), or about local text structure (e.g., signaling words). For lesson goals, this applied if students were supposed to *know* – not *do* – something (e.g., *You know which strategies can be applied before reading*). We also counted how many assignments related to global or local text structure, or genre.

For the analysis of procedural knowledge, we coded the presence of information on how to apply strategies or knowledge on text structures (e.g., *how to scan a text*). For lesson goals, this applied if students were supposed to *know how to do something* (e.g., *You know how to find a main idea*). We also coded the

type of strategy (before, during, or after reading), counted how many student assignments were provided, and whether assignments were aligned with lesson goals and theory (e.g., lesson goal on summarization, assignment asking students to select main ideas).

For the analysis of conditional knowledge, we coded whether information was provided on how to apply reading strategies in other contexts, whether students could plan and evaluate their strategy use, and if there was attention for transfer (e.g., applying strategies in content-area texts). Lesson goals were classified as conditional if students were supposed to learn *when* or *why* strategies work (e.g., *You know when you use a Venn diagram to summarize a text*). Per lesson, we counted how many questions related to conscious strategy planning and evaluation, transfer, and explicit conditional knowledge (e.g., *What type of text would you summarize with a timeline?*).

Coding of instructional features

We also coded several instructional features, as these reflect the intended behaviors, strategies, and interactions that can be expected from teachers and students (Century et al., 2010). The GRR model was used as a framework to compare the textbooks in the sequencing of instructional activities, by coding per lesson in a dichotomous variable the presence of the following components:

- Explicit teacher-led instruction and/or teacher modeling ('I do it').
- Interactive class discussions and/or guided practice ('We do it').
- Collaborative student activities with true shared responsibility and/or working parallel in pairs or small groups ('You do it together').
- Independent practice such as question answering ('You do it alone').

The reliability of the materials analysis was increased by discussing difficult cases during meetings with all authors (Lincoln & Guba, 1985). A subsample of 16 lessons (two per textbook) were also coded by a second coder, resulting in sufficient interrater agreement percentages (Miles et al., 2013). For coding of variables related to declarative, procedural, and conditional knowledge, the initial mean interrater agreement percentages were 87% (min = 69%, max = 100%), 79% (min = 69%, max = 100%), and 91% (min = 75%, max = 100), and for instructional

aspects 84% (min = 64%, max = 100%). Most aspects resulted in high agreement percentages, and for the other aspects, minor differences were resolved in discussion between the two raters.

Interviews

In order to gain insight into teachers' enactment and appreciation of the textbooks, 29 semi-structured interviews were conducted with mainly fourth and fifth grade teachers (86% female, $M_{\text{experience}}=16.4$ years, $M_{\text{age}}=39.7$ years), who had been using their textbook for at least two years ($M = 4.3$ years). Teachers were recruited from 29 schools throughout the Netherlands. The distribution in denominations (9 Catholic, 14 Public, 6 Protestant) and demography (eight different provinces; 50% of schools in large towns) was representative for the Netherlands.

Interview questions covered the same topics as the materials analysis: learning goals, theory and assignments, transfer, and instructional approach (i.e., teacher-led instruction, student activities, sequencing). For each topic, teachers were asked about their actual teaching practice (e.g., for transfer, 'Do you strive to integrate reading comprehension lessons with other subjects? How exactly?'), and about their satisfaction with teaching materials with respect to this topic ('To what extent do the teaching materials support you in doing so?'). All interviews were audio-taped and transcribed with permission of the participating teachers. Verbal data were then inspected for emerging patterns and compared to findings from the materials analysis and lesson observations (Lincoln & Guba, 1985).

Lesson observations

With permission of school leaders and teachers, reading comprehension lessons were observed by one of three observers. We strived for three lesson observations per teaching method, and six for *Nieuwsbegrip*, as this teaching method is strongly in vogue in Dutch schools. Due to unforeseen COVID-19 restrictions, we completed only eleven observations (Table 1). We recruited a different group of teachers for the lesson observations to increase our sample size; only one of the observed teachers also participated in an interview. The participating teachers (83% female, $M_{\text{experience}} = 9.3$ years) taught in grades 4 and 5 or in mixed grades (4-6). The schools were located in two central provinces (36% schools in villages

or small towns), and most schools were religiously affiliated (3 Catholic, 2 Protestant, 2 public).

Observers coded whether certain predetermined features of the lesson content occurred (e.g., mentioning lesson goals, discussing vocabulary, text structure and/or strategies), and how much time was devoted to different lesson phases (e.g., teacher-led instruction, individual question answering) and other instructional aspects (e.g., providing product or process feedback). Due to strict ethics regulations, lessons could not be videotaped, but the validity of the observation scheme was increased by formulating concrete events that were evaluated on occurrence and duration, and by leaving room for a general impression and for remarks, so that difficult decisions could be discussed afterwards (Miles et al., 2013).

3.4 Results

Declarative knowledge

Implemented Curriculum

All textbooks pay relatively little attention to declarative knowledge (Table 2). In 19% of the analyzed lessons, a declarative knowledge goal is formulated ('After this lesson you know that...'). In lesson goals and theory, the main focus is typically not on declarative strategy knowledge, but on strategy application. Six textbooks provide review lessons in which all reading strategies are repeated, but these lessons consist of extensive practice, without explicit strategy instruction. Declarative knowledge is often restricted to mentioning that students can consult the appendix for an overview of all strategies while working on questions. *Lezen in Beeld* pays slightly more attention to declarative knowledge, and also asks students to memorize six different reading strategies.

Textbooks differ from each other most notably in their text selection. In most lessons, *Kidsweek* (KW), *Grip op Lezen* (GL), and *Lezen in Beeld* (LB) offer a relatively rich variety of texts (i.e., varying genres and lengths). *Nieuwsbegrip* (NB) and *Leeslink* (LL) put emphasis on news articles and provide weekly updated texts. The niche method *Bliksem* (BL) focuses primarily on children's literature, and *Lezen is Weten* (LW) prompts teachers to make use of expository texts. However, the

attention devoted to declarative knowledge about genres and text structures is quite limited across textbooks. One textbook explicitly lists knowledge on various genres and text structures among its global learning objectives.

TABLE 2

Materials analysis: Overview of declarative knowledge results

Unit of analysis		Yes	%	Example
Goals	Among general objectives	1		<i>Students learn the purpose, structure, and lay-out of many types of texts.</i>
	Declarative lesson goals	5	19	<i>You know the purposes of advertising.</i>
Theory	Explicit review of all strategies (≥ 4)	5	13	<i>You learn to apply all six key strategies</i>
	Text structure	3	10	<i>There are fixed structures, like question-answer, cause-effect, then-now.</i>
	Signaling words	6	13	<i>'So' is a signaling word for conclusion.</i>
	Basic division intro-body-conclusion	4	5	<i>A text has an introduction, a body and a conclusion. Paragraphs have white lines.</i>
Exercises	Genre, local or global text structure	8	44	<i>What are the features of a report? What is the structure of paragraph 1?</i>

Note. Yes: present in number of textbooks; %: percentage of all analyzed lessons.

Concerning text structures, 13% of all analyzed lessons pay some attention to signaling words, 5% to the introduction-body-conclusion division, and 10% of the lessons mention text structures (e.g., chronology, problem-solution, story grammar), although this only happens in three textbooks. Most declarative knowledge on text structures is restricted to definitions, as in (1) and (2).

1) Cause-effect signaling words tell what causes something (e.g., *therefore*, *as a result*). (GL-4)

2) A text has an introduction, a body, and a conclusion. (TV-4)

Some textbooks provide more elaborate but still relatively abstract explanations. For example, different types of signaling words are mentioned without explaining in what structures or genres students can expect these signaling words, and how they can facilitate comprehension. Only *Leeslink* explains not only which signaling words figure in chronological texts, but also *why* this is useful to know:

3) A text does not always discuss events in their original order. If you make a list of all events ordered in time, you will understand the text better. Signaling words can help you to do this. (LL-5)

While textbooks offer limited information on text structures, a relatively large proportion of student assignments is devoted to text structure (44%). These questions prompt students to recognize genres, structures, and signaling words (4), to complete graphic organizers, or to locate the introduction, body, and conclusion of texts (5).

4) *He hasn't been there for years, ____ he notices many changes in his country.*
Fill in: A. Because B. Therefore C. So that D. Hereby (GL-4)

5) What parts of the text are introduction, body, and conclusion?
The introduction is in paragraph _____. The body is in paragraph _____ to _____. The conclusion is in paragraph _____. (TV-4)

Enacted Curriculum

Declarative knowledge goals are lacking in all of the observed lessons. Instead, most teachers closely follow the guidelines in the textbook, except for the fact that they often add activities related to the topic of the specific text (e.g., video clip on ocean pollution). Four teachers (36%) briefly discuss aspects related to genre or local text structure while providing feedback on assignments (e.g.,

highlighting a signaling word), but they never mention global text structures (e.g., cause-effect).

A similar picture arises from the interviews. The teaching of reading strategies is recognized as important content, while none of the teachers spontaneously mention text structure or genre knowledge as important content. Three teachers even intentionally skip declarative knowledge on genres and signaling words, because they do not feel that it matters for text comprehension, or because they experience difficulties explaining topics related to text structure knowledge (6).

6) I find it very hard to explain signaling words to my students; in fact, I find it even too hard. (...) To be honest, I don't understand why my students should know about signaling words. I don't think it is important. They simply need to understand the text! (TV-teacher)

By contrast, almost one third of the teachers (31%) mention that their textbook does not provide sufficient opportunities to teach about topics related to text structure (e.g., summarization, signaling words, referential coherence), and that they experience difficulties teaching it without guidance of their textbook (7). Some teachers (14%) remediate these problems by giving their students extra texts with questions focused on local text structure (i.e., signaling words and referential coherence); to a much lesser extent by providing additional instruction themselves.

7) There is little on summarization. (...) My students also struggle with signaling words and referential pronouns. There are too few materials about it. (LB-teacher)

Still, most interviewed teachers (76%) do not adapt the curricular content, but try to enact all lessons with dutiful fidelity. In fact, for most teachers, the central consideration in their choice or appreciation of a textbook is not the curricular quality (14%), but rather whether the texts will capture students' interest and keep them motivated to read (34%), and whether the textbook has an attractive layout (21%). Approximately one third of the teachers (32%) explicitly mention that

they are unsure about the curricular structure. Some teachers simply trust their textbook (8), others feel less positive about this, even though they don't make changes (9).

8) I don't see a clear structure [in the curriculum], but I believe that what the textbook offers is what children need to know. (TV-teacher)

9) We want to let go of it [the strict curriculum], and be more adaptive to children's needs. But we cannot do that as long as we don't know what is needed for good reading comprehension. (LL-teacher)

Procedural knowledge

Implemented Curriculum

In all textbooks, emphasis is put on procedural knowledge, especially on practicing reading strategies (Table 3). All brochures mention a fluent use of strategies as their main objective. In all textbooks, a fixed set of reading strategies dictates the curricular structure and lesson goals.

Procedural lesson goals appear in 91% of the analyzed lessons ('After this lesson you are able to...'), most often as the only goal of that lesson (68%). They are evenly distributed over the phases of the reading process: before (31%), during (32%), or after (37%) reading. In some textbooks, the wording of lesson goals focuses more on the central activity than on the procedural knowledge to be acquired (10), or foregrounds text content (11).

10) In this lesson you create a flow chart of the text. (TV-5)

11) You will learn about an owl who learnt to fly with donor feathers, and how making use of prior knowledge can help you. (KW-4)

TABLE 3
Materials analysis: Overview of procedural knowledge results

Unit of analysis		Yes	M / %	Example
Goals	Among general objectives	8		<i>Students learn how to use each of the seven evidence-based reading strategies.</i>
	Procedural lesson goals	8	91%	<i>You create a flow chart; You know that you should scan a text before reading.</i>
Theory	Strategies before reading	8	31%	<i>First, figure out why you want to read the text, e.g., to learn about a topic.</i>
	Strategies during reading	7	32%	<i>If you don't know the meaning of a word, read the text again or look at the pictures. Or you can use a dictionary.</i>
	Strategies after reading	8	37%	<i>In an arrow scheme you can write down the main idea and all the subtopics.</i>
Exercises	Mean number of exercises	8	11.3	
	Related to theory/goal	8	59%	

Note. Yes: present in number of textbooks; %: percentage of all analyzed lessons; *M*: mean number in all analyzed lessons.

Even though most textbooks lack content-related lesson goals (11), they still put an emphasis on text content. For example, 16 lessons (20%) prescribe a lesson start with video clips and/or extensive discussions about the text topic, without linking it to the lesson goal (12). As a result, no or much less instructional time is left for actual strategy instruction.

12) Does anyone play soccer? What is so fun about playing soccer? What is your favorite club? What is your parents' favorite club? Why? (...) (TV-5)

The emphasis on procedural knowledge is also reflected in theory and exercises: in 80% of the analyzed lessons, the theory and instruction concern procedural knowledge and describes how certain reading strategies should be applied. In comparison to the other textbooks, *Kidsweek* provides quite elaborate procedural knowledge about the different reading strategies, often with step-by-step instructions. Except for the two niche methods, textbooks expect students to answer a large number of questions about the text each lesson ($M_{\text{overall}} = 11.3$, $M_{\text{niche methods excluded}} = 15.1$, $\text{max} = 31$). On average, 59% of the tasks or questions per lesson are directly related to the lesson goals and theory.

Enacted Curriculum

Most interviewed teachers (76%) believe that using reading strategies correctly is the main goal of comprehension instruction, and that this is also the primary objective of their textbook and the focus of their lessons. Several teachers (28%) also consider vocabulary growth and acquiring world knowledge as important objectives: their students should have learned the text content upon lesson completion (13). To this end, about half of the teachers (48%) say that they often (re)read the text, and/or extensively discuss students' personal experiences and prior knowledge (38%), or even answer all questions together (17%).

13) Of course, there are those reading strategies, and we do practice them, but it should all be about the text. My lesson goal is always something like: 'I know why a boat sank in the Mediterranean Sea.' That is more fun. (KW-teacher)

Three teachers mention as a side-effect that their students often remember the text topic of the previous lesson, but not the reading-related learning goals. One teacher attributes this to poor alignment of lesson goals, theory, and exercises (14).

14) You have to introduce the lesson goal first, then you discuss the text, and start answering questions. But only halfway the lesson, you get to the questions that relate to the goal (...). They give them almost no opportunity for practice. (TV-teacher)

Lesson observations show that six teachers (55%) do not mention the lesson goal, even though their textbooks provided explicit lesson goals, and only three teachers (27%) actually provide explicit instruction related to the lesson goal. Instead, most teachers (82%) mainly focus on activating prior knowledge on the text topic and building vocabulary. To this end, three teachers (27%) show video clips, five teachers (45%) create mind maps. Seven teachers (64%) also extensively discuss students' experiences or knowledge on the topic. Three teachers (27%) explicitly add their own content-related lesson goal (e.g., 'By the end of this lesson, you can mention the advantages of 3D printers'). Also, strategy practice through questions is a key feature: in all of the observed lessons, most lesson time (42%) is focused on individual question answering.

Most interviewed teachers (69%) display a positive attitude towards fostering procedural knowledge through extensive practice, and seem satisfied with their textbook's focus on comprehension questions. Two teachers (17%) answer almost all questions together with their students; during the observed lessons, even more teachers (36%; four teachers) appear to do this. The value teachers attach to question answering might be one of the reasons why they closely follow their textbooks (15). Four teachers even adapt their curriculum by providing additional texts with traditional question answering (16).

15) I never use other texts, because then I would have to make up the questions myself (...). The questions help my students to think about what they are doing. (LB-teacher)

16) We were so busy practicing those reading strategies, that my students no longer really read texts. So I select texts from another quite old-fashioned book with simple question-and-answer tasks. They learn so much more now! (LL-teacher)

At the same time, various teachers (34%) complain that their students rush through the extensive number of exercises, or that students get bored, especially while working on their own (17). Some of the interviewed teachers (27%) also regret that they often lack time to discuss answers and provide feedback.

17) They find texts and questions too long. Children are too lazy to finish reading the whole text. To me it seems that they are lazy and bored. So in fact, they stop focusing on comprehension during comprehension lessons! (LB-teacher)

Nevertheless, question answering is the main focus in most of the observed lessons (73%), almost as a goal in itself, as shown in a quote by one of the teachers (18). Students' answers to comprehension questions are often only briefly discussed with little attention for the underlying process of text comprehension (64%). Only two teachers (18%) exploit this as an opportunity to provide additional instruction.

18) Even if you have to read the text and the question ten times, you just have to do it. You read the text over and over again, until you know the right answer. (NB-teacher)

Irrespective of the textbook they use, interviewed teachers (34%) feel obliged to have their students practice with question types that typically appear in Cito tests: high-stakes national assessments of reading comprehension. In one of the observed lessons, all instruction and practice focuses on answering Cito multiple-choice questions. For many interviewed teachers, Cito tests have become the gold standard; teachers criticize the quality of their textbook for not being sufficiently aligned with the types of questions and texts in Cito tests.

Conditional knowledge

Implemented Curriculum

It is important that students learn to consciously plan, regulate and evaluate their use of reading strategies. Conditional knowledge is needed for this skill to develop, but the teaching materials pay little attention to this type of knowledge and hardly explain under what circumstances a certain strategy is most powerful (Table 4). Only 5% of the analyzed lessons contain an explicit conditional knowledge goal ('After this lesson you know when/why...'). In 39% of the lessons, textbooks provide some theory about why or when a reading strategy is useful, but this explanation mainly concerns a general statement about the usefulness of a strategy, such as 'This will help you to understand the text.'

Students are almost never at liberty to choose and motivate which reading strategy they want to apply, because exercises dictate students what to do – as in (19) – or because most lessons focus on one strategy only. Even in so-called integration lessons in which students are supposed to review multiple reading strategies, they are almost never asked to plan and evaluate their own strategy use.

19) Read the title of the text. Look at the pictures. What do you think is the topic of the text? About _____. (LL-4).

A positive exception to this pattern is the niche method *Bliksem*. This teaching method focuses more on conditional knowledge than the other textbooks, by having children apply multiple reading strategies on self-selected text passages, and evaluate their usefulness. Only six percent of all analyzed questions tap into conditional knowledge ($M = 0.67$ per lesson), by asking students to evaluate if a certain strategy matched their reading goals (20), or – to a lesser extent – plan their strategy use (21). As most of these questions are part of independent practice, students might not receive feedback on their response.

20) How well did you understand the text? What would you do differently next time? A. Reread parts B. Read slower C. Take notes D. Other: (NB-5)

21) When would you make a summary: while reading a book for pleasure, or while studying for a history test? Why? (BL-5)

TABLE 4

Materials analysis: Overview of conditional knowledge results

Unit of analysis		Yes	M / %	Example
Goals	Conditional lesson goals	2	5%	<i>You understand why you should make use of prior knowledge while reading</i>
Theory	Conditional knowledge	7	39%	<i>This strategy will help you understand a text better</i>
Exercises	Mean number of exercises on planning and evaluating strategy use	6	0.69	<i>How well did you understand the text? What would you do differently?</i>
	Mean number of transfer exercises	6	0.68	<i>Write a short text about animals. Do not use referential pronouns. Now read your story. Do you like how it is written?</i>
	Suggestions for transfer in teacher manual	6	53%	<i>When you read other texts, ask students if these are about facts or opinions.</i>
	Transfer tasks aligned with goals	6	35%	<i>Text about art: "Close your eyes. How do you feel? Express it in a painting"</i>

Note. Yes: present in number of textbooks; %: percentage of all analyzed lessons; M: mean number in all analyzed lessons.

Although most information brochures mention that transfer is of paramount importance, lesson plans offer little opportunities for students to apply reading strategies in other contexts. Attention to transfer often consists of suggestions for teachers to model a specific strategy during other lessons (53%). Some textbooks contain tasks other than comprehension questions (e.g., aimed at writing, speaking, or reading content-area texts) that might promote transfer, but such transfer tasks are limited ($M = 0.68$), and are often presented as optional assignments for high-ability students.

These transfer tasks typically focus on writing and speaking (43%), the reading of content-area texts (35%), or involve creative exercises, such as making a drawing related to the text topic (22%). A closer inspection shows that 65% of these tasks are not related to the learning goal, but to the text topic and cannot be classified as true transfer tasks. For instance, a lesson with a text on shipping contains a task in which students write a captain logbook and make a drawing.

Enacted Curriculum

More than a quarter of the interviewed teachers (28%) express doubts related to the development of conditional knowledge. For instance, they wonder if textbooks provide students with the knowledge needed to plan and evaluate their reading process (22 and 23). At the same time, some teachers (21%) are negative about lessons in which students are asked to apply multiple strategies more independently. They often skip these lessons, as they are afraid that their students are ill-prepared and would get overwhelmed. By contrast, the teachers who use the niche method *Bliksem* are enthusiastic about the way in which their textbook promotes metacognitive skills.

22) In most textbooks, the focus is too much on writing down answers to a series of questions, and too little on critical thinking. (BL-teacher)

23) There is little room for students' own initiative. For example, supposedly difficult vocabulary is already underlined and explained. They do not let students come with their own questions and solutions. (KW-teacher)

This lack of conditional knowledge is also apparent in the observed lessons: only one teacher explains when and why it is useful to summarize a text. None of the other teachers explicitly discuss when or why the lesson goal or theory of that lesson could be useful. For instance, in one lesson, the teacher provides excellent instruction on how to use a dictionary, without ever mentioning in what situations a dictionary search is useful, so that students might end up with the wrong idea that you have to look up every single difficult word you encounter in a text, while less effortful strategies could be used as well. Most teachers follow their textbook and the clear-cut student exercises, but two teachers (18%) provide their students with more freedom to apply multiple strategies while reading, by asking them to circle and search the meaning of unknown words, and to highlight key information.

Most teachers find it important that their students apply their knowledge of reading comprehension in contexts outside reading class. Therefore, teachers try to review some reading strategies during content-area (58%) and language lessons (31%), and some during arithmetic (10%) or writing instruction (7%). This attention to transfer is often spontaneous and typically restricted to simply mentioning an appropriate reading strategy or pointing at a list of strategies on a poster, instead of modeling a specific strategy in a new context. Teachers (38%) quite often mention that they skip the transfer exercises provided by their textbook, because they consider them as optional, or feel not confident to teach them (24).

24) We find it difficult to apply strategies during content-area lessons, because you should select texts yourselves and look for a suitable strategy. We tried to do it for geography, but we don't feel confident without a manual. (BL-teacher)

The instructional approach

Implemented curriculum

Figure 2 shows the sequencing of instruction and activities that is prescribed in the textbooks, showing large variability across textbooks: some follow the full GRR

model, some only for low-ability students, and other textbooks show no gradual release pattern. Two textbooks offer fully teacher-centered explicit strategy instruction, while the other textbooks provide instruction through modeling or interactive discussions.

In the textbooks with modeling, the directions for teachers are diverse: some manuals briefly describe what teachers should model, whereas four textbooks regularly provide complete scripts that literally dictate what teachers should think aloud. However, these scripts do not always respect the basic rules for effective modeling. For example, interactive questions are added in the middle of a think-aloud which does not respect students' primary role as observers during modeling, or the script leaves out the verbalization of the exact thinking processes involved (25): what did the teacher read in the surrounding text that made him realize the meaning of the word?

25) [Modeling script] I see a difficult word: gestation period. Let's see if I can find out the meaning of the word by reading in the surrounding text. Yes! Now I don't need to check the meaning in my dictionary anymore.
(TV-4)

Although a moment of abstraction is important for effective modeling, none of the textbooks prescribes reflective post-modeling activities that help students reflect on their teachers' reading approach and on the effect of the employed reading strategies. One textbook suggests that students should engage in modeling under supervision of their teacher, but in most textbooks, the lesson simply continues with a new kind of activity. Two textbooks systematically suggest guided practice before students start working on their own or in groups, but guided practice is often lacking, or only prescribed for low-ability students.

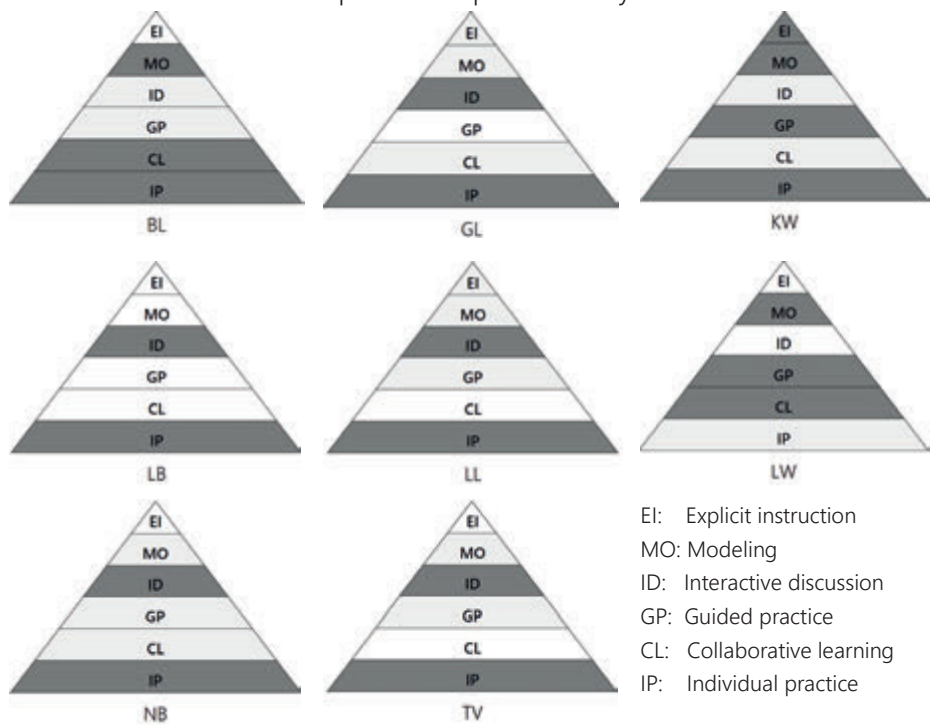
By contrast, all teaching materials provide ample opportunities for individual student activities, but collaborative activities – especially in small groups – are much less structurally part of the instructional model. In most cases, collaborative activities consist of small-group question answering instead of structured collaborative learning activities (26).

26) Have students make exercises 1 to 3 in pairs or groups of three. They can discuss quietly.

Only two textbooks structurally integrate collaborative learning activities in which the task design goes beyond simply answering questions in pairs or small groups.

FIGURE 2

Presence of instructional components as prescribed by the GRR model



Note. Dark grey: often (≥ 80% of lessons), light grey: sometimes, and white: never.

Enacted curriculum

Although most teachers taught *Nieuwsbegrip* instruction lessons (see Methods) – a textbook that incorporates most phases of the GRR model (Figure 2) – lesson observations revealed considerable differences in teachers’ enactment. At the extremes were one teacher who followed a full GRR, and two teachers who only provided explicit instruction, directly followed by independent practice. On average, approximately 23% of the total lesson duration was devoted to teacher-led instruction – mainly interactive discussions between teacher and students, and 42% was devoted to individual practice (Table 5). Teachers spent more time on guided practice (i.e., answering questions together) than was generally prescribed by their textbook. Only one teacher guided students in how to collaborate on a given assignment, instead of stating to simply ‘work in pairs’. Interestingly, two teachers (18%) explicitly discouraged all collaboration (27):

27) I want that you make all the questions on your own, not together. You need to be able to do it all by yourself. (NB-teacher)

TABLE 5

Observed lesson time (in minutes) allocated to each instructional phase

Phase	Number of teachers	Time (SD)
Explicit (interactive) instruction	10	10.60 (8.03)
Modeling	3	1.60 (1.99)
Guided practice	7	10.80 (12.15)
Collaboration or working in pairs	4	3.00 (4.52)
Individual practice	8	19.80 (14.12)

Teacher interviews revealed that teachers make fewer adaptations in the instructional approach when their textbooks have elaborate guidelines, and that they seem more inclined to make adaptations when textbooks provide less scripted lessons. For instance, one teacher who uses *Lezen in Beeld* – only prescribing interactive discussion and individual question answering – mentions how she provides modeling, and how she has even started to promote peer modeling, so that students can learn from each other (28).

28) I try to let my students model as well, so they will start modeling as if they are the teacher themselves. To show their thoughts to other students.
(LB-teacher)

Although teachers are generally positive about the availability of modeling scripts, they also report that they sometimes rely too heavily on the scripts, or even read out the scripts to their class, especially if they lack preparation time or feel insecure (29). As they obediently follow the scripts, teachers also copy flaws in the materials and show the same tendency to ask interactive questions during modeling.

29) Then you just open your binder and simply read out what is written. Some kids say: 'Miss, are you just reading out what is written there?'
(BL-teacher)

Only two teachers (7%) report that they add a self-initiated moment of reflection after modeling. This resonates with findings from the lesson observations: in half of the modeling instances, teachers did not really think aloud, but engaged in interactive reading, and in only one of the observed lessons (9%), the teacher discussed her strategy application after modeling.

Some teachers (28%) regret the fact that their textbook does not provide support for collaborative learning. They have students collaborate on their own initiative. Other teachers (31%) say they would like to integrate collaborative learning, but that they cannot have students collaborate, because of organizational challenges (e.g., mixed-grade teaching), ability differences, or because they feel they have to monitor students' individual achievements. Various teachers report that in the current form of 'working in pairs', collaboration goes no further than simply copying each other's answers (30).

30) There always is the risk that one student copies the answer of the other student, or that they try to finish the questions as quickly as possible and rush through it all. (TV-teacher)

3.5 Conclusion and Discussion

An analysis of what textbooks offer and what teachers teach during reading comprehension lessons provides valuable insight into the quality of reading curricula. This mixed-methods analysis reveals various aspects of reading strategy and text structure instruction that could be improved in the Dutch curriculum.

Of course, our study has its limitations. Although we strived for data sources equally representing all eight different textbooks, our observational data overrepresent *Nieuwsbegrip* due to COVID-19 restrictions, which might limit the generalizability of our observational findings. Still, it is important to note that the observations often confirmed findings from the interviews and materials analysis, in which all textbooks were equally represented. Despite individual differences between teachers, we found robust patterns that were confirmed by methodological triangulation. Furthermore, the findings of this study seem consistent with international studies on textbook quality (Beerwinkle et al., 2018; 2021; Seifert, 2021; Wijekumar et al., 2021; Zhang et al., 2021).

Another issue was the fact that, although the interrater reliability for instructional aspects was satisfactory, it was not always easy to code instructional activities. Textbooks did not always provide unambiguous guidance, or sometimes the design of lesson activities made it challenging to distinguish true collaborative learning (e.g., Johnson & Johnson, 1984; 2017) from tasks in which students simply had to answer comprehension questions in pairs.

The main findings are as follows. First, with regard to curricular content, we found that the implemented and enacted curriculum look very much alike. Most teachers follow the curricular content linearly, despite their complaints about the clarity or usefulness of the curriculum, thereby copying the poor alignment between lesson goals, theory and exercises in their classroom. This has been reported in previous studies as well: teachers with limited pedagogical content knowledge often follow their materials with 'dutiful fidelity' (Beerwinkle et al., 2018; Dewitz & Jones, 2013; Valencia et al., 2006; Wijekumar et al., 2019a) and are at risk to copy flaws (Aarnoutse, 1990). By contrast, self-confident and knowledgeable teachers critically examine, flexibly adapt, and actively evaluate the right sequence of instruction, lesson plans, and activities (Jitendra et al., 2001; Penuel et al., 2014; Piasta et al., 2009). Only a few teachers explicitly mention this problem, wishing

they had more knowledge to flexibly adapt their instruction to students' needs and become less dependent on their textbooks.

By contrast, more variation between the implemented and enacted curriculum can be observed with regard to the instructional approach: textbooks offer a wide array of activities – that unfortunately does not always reflect a gradual release of responsibility (GRR) – but teachers do not necessarily copy the textbook approach. They often put their own emphasis in their enactment of activities. On a positive note, this experienced freedom could allow teachers to provide a GRR in flexible and authentic ways, depending on students' needs, and avoid implementation rigidity (McVee et al., 2018; Webb et al., 2019).

However, the resulting approach often does not reflect a GRR: teachers showed fewer instances of modeling and more individual practice than their textbooks offered, so that the opportunities for scaffolding and metacognitive knowledge development are limited. This problem is amplified by the fact that textbooks not always provide a GRR and lack cues to inform teachers about the right amount of scaffolding in various phases. Yet, without guidance, teachers with limited knowledge might expect their students too quickly to implement reading strategies independently (Almasi & Fullerton, 2012; Shanahan et al., 2019).

Second, although reading researchers have advocated a balanced focus on declarative, procedural, and conditional knowledge to teach strategies, the current curriculum displays an imbalance between these types of knowledge. Due to the emphasis on procedural knowledge (i.e., procedural lesson goals, extensive practice of strategies in fixed tasks, texts-with-questions approach) in both textbooks and classroom practice, reading strategies have become a goal in itself rather than a means to an end. This can be detrimental to students' text comprehension skills (e.g., Sinatra et al., 2002; Wilkinson & Son, 2011). That is, for effective strategy instruction, it is fundamental to pay attention to the underlying process of goal-directed decision making on the right reading approach in a given situation (Afflerbach et al., 2008; Alexander, 2018; Paris et al., 1983).

To some extent, practicing reading strategies is good; it can turn these strategies eventually into effortless skills that can flexibly be used to support comprehension (e.g., Afflerbach et al., 2008; Alexander, 2018). For teachers, comprehension questions are easy to implement and they facilitate student monitoring. However, this text-question-answer model has limited desirable

effects on reading comprehension (Rooijackers et al., 2020; 2021). Moreover, many questions in our sample turned out to be unrelated to lesson goals and theory, so that students lack the necessary knowledge for task execution, and ignore the purpose of these questions. Several teachers noticed negative effects of this lack of alignment: students rush through the numerous questions and often remember the text topic, but not the lesson goal, which echoes earlier criticism on the Dutch reading curriculum (Berends, 2011; Wieberdink & Kuster, 2015).

Unfortunately, the observed teachers seemed to copy the lack of alignment in their classrooms: they often did not explicate the learning goals – even though their textbooks provided them – and strongly focused on text content and right answers. Although some teachers criticize the text-question-answer model, it still dominates reading lessons, sometimes to the point that it actually undermines a gradual release of responsibility. That is, even though students could learn from each other through collaborative learning activities– such as reciprocal teaching, peer modeling, and peer-assisted learning strategies (Fuchs et al., 1997; 2021; Hacker & Tenent, 2002; Keehnen et al., 2015; Palinscar & Brown, 1984; Rosenshine & Meister, 1994) – most textbooks only provide comprehension questions that can be answered in pairs. This does not create optimal conditions for collaborative learning (see Johnson & Johnson, 1984; 2017) and focuses students' attention on the reading product. Even if collaborative activities are provided, several teachers still assign them as individual activities, which is motivated by practical reasons (e.g., classroom management), as well as by teachers' desire to prepare their students for high-stakes reading tests.

To some extent, the overemphasis on individual questions might be understood as a negative washback effect of the Dutch testing culture (Bartels et al., 2012; Scheerens, 2017). Several teachers criticized their textbook for not being sufficiently aligned with high-stakes reading tests, and/or even added test taking practice. This illustrates how teaching has become at the service of testing instead of the other way around. This might undermine the instructional time devoted to higher-order thinking skills (Cheng & Curtis, 2004; Hamp-Lyons, 1997). Rouffet et al. (2022) reached a similar conclusion with respect to foreign language teaching: high-stakes tests strongly affected how teachers shaped their teaching.

The third main problem that was uncovered is that the curriculum lacks systematic attention for conditional knowledge development. Students should

receive instruction on the nature, function, importance, and application of reading strategies (Okkinga et al., 2018b), and should have ample opportunities to independently choose and evaluate their use of reading strategies, in increasingly complex texts and genres (Afflerbach et al., 2008; Alexander, 2018; Simpson & Nist, 2000). However, Dutch students are almost never asked to plan and evaluate their strategy use, as lessons often focus on one or two strategies, and questions specifically dictate what strategy should be applied next. If conditional knowledge is offered, it concerns abstract notions of *why* a strategy works, but not *when* or *where* it should be applied. This is copied by teachers: they hardly pay attention to conditional aspects, and tend to skip lessons in which students have to apply strategies independently, because they consider them optional or too difficult.

This problem is even more amplified due to the fact that, in the Netherlands, the different skills underlying reading comprehension are taught in isolation, and reading comprehension is taught as a separate subject. This is often perceived as demotivating (Berends, 2011), and makes the transfer of reading strategies to other subjects more challenging (van Gelderen & van Schooten, 2011). This has also been noticed in other countries: reading skills are often taught in isolation (Wijekumar et al., 2019) with very limited attention to conditional knowledge (Snow, 2002), even though developing readers should learn to be metacognitive, by deciding which strategies work best in which context and by learning to coordinate multiple strategies (Afflerbach et al., 2008; Reutzel et al., 2005; Rogiers et al., 2019; Simpson & Nist, 2000; Soto et al., 2019).

The declarative knowledge component is also shallow in Dutch textbooks, especially when it comes to text structures, even though explicit text structure instruction relates to better comprehension, recall, and summarization skills in primary education (*Chapter 2*) and beyond (Hebert et al., 2016; Pyle et al., 2017). A similar lack of attention to text structures in textbooks has been reported in other countries (e.g., Austria: Seifert, 2021; China: Zhang et al. 2021; Malta: Aguis & Zammit, 2021; US: Beerwinkle et al., 2018; 2021; Wijekumar et al., 2019a; 2021). A striking similarity with these studies is the fact that if text structure is discussed, it is often not made explicit to students why and how knowledge of text structure is needed to gain a deeper understanding of the text, which underscores again the underrepresentation of conditional knowledge.

Dutch teachers seem relatively unaware of this gap and do not add their own materials on text structures, even if this might be warranted to support their students' comprehension. In fact, most teachers do not feel that text structure is an important topic, although some teachers complain that their students struggle with questions related to text structure. Teachers often try to fix these problems by providing their students with additional practice – not explanation, which suggests that teachers themselves also lack the knowledge and skills to provide high quality instruction on text structure (see also Beerwinkle et al., 2018; Seifert, 2021; Wijekumar et al., 2019a; 2021). Moreover, the fact that teachers simply provide more individual practice instead of activities with more opportunities for scaffolding shows that teaching manuals should provide better cues for teachers to determine the right amount of scaffolding, so that a gradual release of responsibility will be restored (McVee et al., 2018; Shanahan et al., 2019).

Not surprisingly, discussions about textbook quality often end with a reference to the Peter Effect: teachers cannot give what they do not have (Applegate & Applegate, 2004; Beerwinkle et al., 2018). Even though we did not assess teachers' general knowledge and beliefs about reading comprehension, teachers' own comments and enactment show how they hardly compensate for suboptimal textbook content. This only underscores why it is crucial that to invest in both the quality of teaching materials for reading comprehension (e.g., Dewitz & Jones, 2013; Pereira & Nicolaas, 2019; Valencia et al., 2006; Van den Branden et al., 2019; Wijekumar et al., 2021), and to invest as well in teachers' pedagogical content knowledge to support their classroom enactment and make teachers less dependent on the quality of their textbooks (Gudmundsdottir & Shulman, 1987; Scheerens & Blömeke, 2016; Schneider et al., 2005; Wijekumar et al., 2019b). For example, in the current Dutch teacher training curriculum, the knowledge provided about text structures is shallow and incoherent (Kooiker-den Boer et al., 2019) and teachers might also need more training in effective pedagogies for strategy instruction (Okkinga et al., 2018b; 2021).

Another interesting area to explore is the development of so-called 'educative curriculum materials' that are designed to simultaneously address both teacher learning and student learning (Ball & Cohen, 1996; Davis & Krajcik, 2005; Davis et al., 2017; Edelson et al., 2021; Haas et al., 2021; Schneider et al., 2005). Educative curriculum materials offer content for teachers beyond the level

suggested for students, and provide training on underlying instructional approaches. For example, with regards to modeling, educative textbooks could provide some background knowledge for teachers on observational learning and effective modeling, as well as improve modeling scripts by clarifying the thinking process and including reflection activities focused on the reading process *after* the modeling phase, instead of prompting teachers to ask interactive questions *during* modeling (see Rijlaarsdam, 2005; Schutz & Rainey, 2019). This could simultaneously strengthen the quality of textbooks as well as teachers' PCK.

In sum, the current study has revealed various knowledge aspects of reading strategy and text structure instruction that could be improved in the current curriculum. Of course, these are only two of the components underlying the complex skill of reading comprehension. Future studies could also examine other aspects of the reading curriculum, or more specifically how teachers try to differentiate their instructional approach. It might also be valuable to examine the moderating impact of teachers' knowledge and beliefs on effective reading comprehension instruction in relation to their classroom practice and textbook use. This might give more insight into what is needed to have knowledgeable and flexible teachers who can compensate for sometimes less optimal materials. After all, reading comprehension is a complex skill to master, and probably even a more complex skill to teach.

CHAPTER 4



Teachers and researchers as co-designers? A design-based research on reading comprehension instruction in primary education

Abstract

As an attempt to bridge the gap between science and educational practice, a design-based research (DBR) was conducted. Four teachers designed and implemented lessons on informational text structures, under close supervision of two researchers. The aim of the study was to gain insight into the viability of the design principles and into the level of support teachers need in order to become effective co-designers. Based on data from lesson artefacts, logbooks, panel interviews, and lesson observations, this chapter presents several difficulties related to the design principles and their implementation in practice. These difficulties were partially due to a tension between two design principles, and to the fact that one design principle needed refinement. However, most difficulties appeared related to teachers' relatively limited pedagogical content knowledge. As a result, the teachers needed a high level of support. Teachers' beliefs and habits also seemed to affect the implementation of design principles, especially when it came to text selection and teachers' views on effective modeling. This study raises questions about the feasibility of equal participation of researchers and teachers at the start of a DBR project, and shows how DBR can contribute to teacher professionalization if researchers provide adequate support throughout the design process.

This chapter has been published as:



Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J. & van den Bergh, H.H. (2019). Teachers and researchers as co-designers? A design-based research on reading comprehension instruction in primary education. *Educational Design Research*, 3(1), 1-23.
<https://doi.org/10.15460/eder.3.1.1399>

4.1 Introduction

For too long, science and education have been two separate worlds: researchers develop scientific knowledge, while practitioners (teachers, policy-makers, educational publishers) hardly translate this knowledge into classroom practice (Broekkamp & van Hout-Wolters, 2007; Brown, 1992; Ormel et al., 2012). Two factors contributing to this research-practice gap are the fact that there is almost no collaboration on equal terms between practitioners and educational researchers, and that practitioners rarely have the time or skills to consult educational research, let alone to translate it into practice. As a result, many parties feel that teaching materials and the curricula of teacher training colleges – and hence teacher knowledge – are seldom research-based (Broekkamp & van Hout-Wolters, 2007).

In the Netherlands, a research-practice gap is also present in the context of reading comprehension instruction (*Chapter 3* - Bogaerds-Hazenberg et al., 2022a). According to the Component Model of Reading Comprehension (CMRC; Aaron et al., 2008), it is highly problematic when teacher knowledge and textbook content are poor, as the instructional context influences students' acquisition of literacy. Unfortunately, curricular materials often follow educational trends and are essentially market driven, while rarely initiating evidence-based principles (Chambliss & Calfee, 1998; Dewitz & Jones, 2013). In addition, teacher knowledge about effective practices for reading instruction appears to be limited (e.g., Koenig, 2018; Kucan et al., 2011; Piper, 2019), especially when it comes to explaining text structure (Beerwinkle et al., 2018; Reutzel et al., 2016) and providing high-quality reading strategy instruction (Okkinga et al., 2018b). Even if teaching materials are evidence-based, this is no guarantee that teachers successfully implement them. For instance, even with adequate curricular products for text structure instruction, teachers do not always teach as intended, because they possess limited knowledge of text structure themselves (Reutzel et al., 2016).

One way to bridge the research-practice gap is to have teachers and researchers collaborate in design-based research (DBR; Broekkamp & van Hout-Wolters, 2007; Ormel et al., 2012; Vanderlinde & van Braak, 2010). Teacher professionalization is often considered to be an important outcome of DBR (Kafyulilo et al., 2016), because the professional interactions between co-designing teachers and experts can influence teachers' knowledge, perspectives, and self-

efficacy (Bell & Gilbert, 1996; Englert & Tarrant, 1993; Voogt et al., 2011). Specifically, teachers' experiences as co-designers can provoke them to rethink their understanding of teaching, and restructure their core ideas, beliefs, and habits (Bannan-Ritland, 2008). Also, when teachers actively design or modify an intervention, this can have a positive impact on its viability and outcomes, as teachers' involvement can make the intervention more fit into the context (Englert & Tarrant, 1993; Harn et al., 2013).

The research focus of a DBR is not on a controlled test of a researcher-designed intervention, but rather on what can be learned from the iterative process of designing and refining a new product or intervention (Barab, 2014). Both teachers and researchers share the responsibility for designing and gaining theoretical insight (Bannan-Ritland, 2008; Englert & Tarrant, 1993). This shifts researchers' focus towards practical design questions, and teachers' focus towards a more theoretical perspective on the problem (Voogt et al., 2011).

For researchers, DBR provides information about the practicality and sustainability of their design principles (McKenney & Reeves, 2018), as the design is repeatedly tested and (re)adjusted within the complex, authentic context of schools (Brown, 1992), and teachers bring their personal expertise and experiences to the design (Hultén & Björkholm, 2016). For practitioners, DBR participation stimulates professionalization (Kafyulilo et al., 2016). The resulting curricular products are also likely to be sustainable, as the shared development of materials increases teachers' sense of ownership (Cviko et al., 2013; Wikeley et al., 2005).

Especially long-term co-design projects create opportunities for teachers to integrate research-based practices into their lessons, and to reconsider their current practice, while it challenges researchers to share power with teachers (Bannan-Ritland, 2008). Instead of a traditional top-down model of knowledge dissemination from research to practice, teachers' involvement leads to a two-way flow of information between teachers and researchers, which encourages all parties to negotiate findings (Nutley et al., 2007; Vanderlinde & van Braak, 2010).

Such a collaborative partnership on equal terms between teachers and researchers is more in accordance with a democratic epistemology in which knowledge is co-constructed in a "reflexive contract" (Bednarz et al., 2012): the theory-driven knowledge, views, and experiences of the researchers are mixed with the context-based experiences, knowledge, and routines of practitioners in a

continuous process of negotiation and reflection. Eventually, this leads not only to democratic knowledge, but also to teacher professionalization (Hultén & Björkholm, 2016). However, the teacher-researcher collaboration may be different in nature and less on equal terms if the knowledge gap is wide, and teachers lack the information that is necessary to reflect and build upon during the DBR.

There may be several threats to a successful collaboration. First, researchers are not used to formulating practice-oriented theories that are fit for the unpredictable realities of classrooms, and are sensitive to teachers' needs (Harn et al., 2013). As a result, the original intent from design principles might get lost when implemented. A close inspection of the implementation can supply insight into the viability of design principles, and provides input to adapt them (Brown, 1992). Second, teachers are not used to translating research into practice (Broekkamp et al., 2007). They may have insufficient content knowledge to successfully implement a design principle, interpreting it incorrectly (Hultén & Björkholm, 2016), or they may not always identify which elements of the intervention are flexible, and which are fundamental and should not be altered (Harn et al., 2013). If such problems arise during the DBR, this will shed light on the support teachers need to successfully implement design principles.

In the current study, we report on how four teachers translated a set of researcher-provided design principles into practice, under extensive supervision of two researchers. The study is focused on text structure instruction in the upper elementary grades. So far, not many DBR studies have been undertaken in the field of literacy instruction (Anderson & Shattuck, 2012), and only seldom do primary school teachers design their own lesson materials for reading comprehension: they follow their teaching materials, even if these are rather weak, and only partially incorporate evidence-based practices (e.g., *Chapter 3*; Dewitz & Jones, 2013). By engaging primary school teachers in a teacher design team (TDT), we aimed at designing text structure lessons, and simultaneously gaining more theoretical insight. Also, we intended the TDT to become more aware of effective practices for reading comprehension instruction. Our research questions are:

- RQ1: How viable are the design principles in practice?
- RQ2: What level and type of support do teachers need to successfully translate the design principles into practice?

4.2 Theoretical Framework: the four design principles

The teachers were asked to implement four design principles (DPs) that were derived from research (Table 1). In this section, we discuss their legitimacy.





DP1: Embed reading instruction in content-area subjects

When students have a purpose for reading and are aware of the relevance of the learning activity, they are more likely to learn from text (Britt et al., 2018; Goldman, 1997; Moss, 2004). Reading comprehension instruction can be made relevant and purposeful by embedding it in the content-area classroom, as it typically provides an authentic context where students have to read vast amounts of expository texts in order to acquire knowledge or complete information-sharing projects (Moss, 2004; Ogle & Blachowicz, 2002; Read et al., 2008). As students are often unfamiliar with the text structures that they encounter, the content-area classroom forms a relevant context for text structure instruction in which students can be provided with authentic projects that require them to summarize or recall information in speaking or writing tasks (Moss, 2004; Read et al., 2008; Reutzel et al., 2016; Williams et al., 2014). During the DBR, the teachers were encouraged to think about engaging tasks that could define a clear reading purpose, so that reading comprehension would be a means to an end.

DP2: Focus on informative text structures

According to the Construction-Integration Model (Kintsch, 1988; 2004; 2013; van Dijk & Kintsch, 1983), successful readers create a coherent mental representation of the text. That is, they make meaningful connections among different parts of the text, and between the text and their own prior knowledge (van den Broek et al., 2005). Text structure can facilitate understanding “by helping the reader to organize concepts based on the explicit or implied relationships that are communicated by the text” (Meyer & Ray, 2011:128). This promotes understanding and improves the recall of text, as text structure helps readers to chunk and organize new information, and provides mnemonic hooks for learning (Anderson & Pearson, 1984; Meyer & Freedle, 1984; Meyer & Rice, 1984; Meyer & Ray, 2011).

TABLE 1
Design principles for teaching text structure

Design principles (DPs)		Teacher role	Learning outcome
DP1 <i>Embed reading instruction in content-area subjects</i>		Define a clear purpose for reading that is close to the content-area curriculum.	Become more proficient in applying reading strategies in authentic, purposeful content-area reading contexts (transfer).
DP2 <i>Focus on informative text structures</i>		Focus students' attention on top-level structure of expository texts and demonstrate how to apply reading strategies to the text structure at hand.	Be able to recognize four expository text structures and apply reading strategies (predicting, questioning and summarizing) in relation to these structures.
DP3 <i>Balance declarative, procedural and conditional knowledge</i>		Work towards goals aimed at knowledge about structures and skills in planning, using, and evaluating reading strategies.	Obtain declarative knowledge about text structures, know when and how to make use of this knowledge while reading (self-regulation), thereby also strengthening reading strategy knowledge.
DP4 <i>Ensure a gradual release of responsibility with an emphasis on modeling</i>		Act as a teacher/model first, and then gradually become a coach. Provide opportunities for reflection, collaborative and individual practice.	See DP2 and DP3.

Informational text structures such as cause-effect and compare-contrast are quite common in the content-area curriculum (Pyle et al., 2017; Williams et al., 2014), which is why text structure instruction is highly recommended by reading researchers (e.g., Meyer et al., 2018). Various meta-analyses have shown the positive effects of text structure instruction on comprehension, especially when multiple text structures are taught (*Chapter 2*; Hebert et al., 2016; Pyle et al., 2017). In the US, educators have strongly recommended text structure instruction, even starting from kindergarten (Duke et al., 2011; Shahanan et al., 2010), but Dutch lesson materials pay little attention to informational text structures (*Chapter 3*).

Therefore, we encouraged the TDT to develop lessons about four informational text structures: cause-effect, problem-solution, compare-contrast, and sequence/chronology. The TDT was specifically asked to (1) focus their text structure instruction on the characteristics of each structure (central questions, cue words), and (2) teach how to apply reading strategies such as predicting, questioning and summarizing specific to the text structure at hand. As teachers often have difficulties with recognizing text structures (Beerwinkle et al., 2018; Reutzel et al., 2016), the TDT received explicit instruction on the four structures.

DP3: Balance declarative, procedural, and conditional knowledge

Teaching reading strategies can promote students' reading comprehension (Bimmel, 2001; Cain, 1999; Droop et al., 2016; Houtveen & van de Grift, 2007; Okkinga et al., 2018a; Palinscar & Brown, 1984; Yuill & Oakhill, 1988). However, students will only become strategic readers if they are able to purposefully apply these strategies, that is, in a flexible, context-sensitive way (Alexander, 2018). Therefore, students need declarative, procedural and conditional knowledge: they need to know (1) which strategies can be applied, (2) how to carry out these strategies as effective as possible, and (3) when and why certain strategies are useful in various types of texts (Duffy, 2002; Duke & Pearson, 2002; Paris et al., 1983; Kostons et al., 2009). Knowing when and why certain reading strategies can be used – conditional knowledge – is crucial for a satisfactory level of self-regulated strategy use (Alexander, 2018; Bimmel & Oostdam, 1999; Malone & Mastropieri, 1992; Nash-Ditzel, 2014). Text structure can provide a useful framework to develop conditional knowledge and can support students in

applying reading strategies in a context-sensitive way. Therefore, the TDT was asked to provide a balance in declarative, procedural and conditional content.

DP4: Gradual Release of Responsibility

The TDT was encouraged to pace instructional activities according to the *Gradual Release of Responsibility Model* (GRR; Fisher & Frey, 2021; Pearson & Gallagher, 1983). According to the GRR model, the responsibility for the learning activities first lies with the teacher, and is then gradually transferred to the students, which is often translated into different lesson phases: explicit instruction and teacher modeling, guided practice, collaborative activities and individual practice.

As many teachers struggle with modeling, and current teaching materials provide little guidance for it (*Chapter 3*; Okkinga et al., 2018b), the TDT received ample instruction on effective modeling. For example, we recommended that after modeling, teachers and students would reflect on what they just observed, because this has a positive impact on student outcomes, at least in the area of writing (Braaksma et al., 2002; Rijlaarsdam et al., 2005). We also explained that students might benefit from comparing models (Gentner & Namy, 1999), and from looking not only at highly competent and self-confident mastery models, but also at less competent coping models who show their errors and hesitations, but gradually improve their performance (Braaksma et al., 2002)

We also explained to the TDT the five essential conditions for effective collaborative learning: positive interdependence, promotive interaction, individual accountability, group processing, and social skills (Johnson & Johnson, 1984; 2017). Two examples of collaborative learning were highlighted: reciprocal peer tutoring (Palinscar & Brown, 1984) and jigsaw activities (Aronson, 1973). In reciprocal peer tutoring, students practice reading strategies in pairs while taking turns as tutee (i.e., read and think aloud) and tutor (i.e., provide support, ask clarification questions) (Palinscar & Brown, 1984). In the jigsaw cooperative learning structure (Aronson, 1973), the reading materials are split into different pieces that are divided among expert groups (AAA, BBB, CCC). Within these expert groups, students try to make sense of their piece of information. Then, mixed groups (ABC) are formed in which each expert student teaches their part to the other group members. This exchange is crucial for the completion of a joint product, just as in a jigsaw each piece is needed (Colosi & Zales, 1998; Doymus, 2007).

4.3 Method

Participants

Over the course of one year, a team of four teachers developed a lesson series for reading comprehension in collaboration with two researchers². The teachers taught at four Dutch primary schools and had on average 19.5 years of teaching experience ($SD = 6.8$). They were purposefully sampled and invited to participate in the DBR, because they were all eager to change their reading instruction. Two teachers (T1 and T2) taught in sixth grade and had some experience with lesson design, and the other two teachers (T3 and T4) taught in fourth and/or fifth grade, and had not developed lessons before, but used teaching materials with a focus on reading strategy practice.

As the design principles were already determined by the researchers, there was some asymmetry in the roles of teachers and researchers at the start of the project. Although DBR would typically imply a reflexive contract in which researchers and teachers co-construct knowledge based on theory, experiences, and continuous reflection (Bednarz et al., 2012), our project gave the primary responsibility to researchers: they formulated the DPs and supervised the teachers during the design process. However, the researchers paid close attention to teachers' reactions to the research problem and DPs, as we wanted teachers to consider themselves problem owners as well. The researchers were also open to teacher feedback. The first author participated as moderator, observant, and primary supervisor to teachers; the second author acted as secondary supervisor and was present during the four teacher meetings. Such an active role of researchers in the design process is not uncommon in small-scale DBR projects with primary school teachers (Ormel et al., 2012).

Design process

Table 2 provides an overview of the design cycles and the data sources. Teachers were invited to two meetings before they started the lesson development. During

² This study was approved by the Faculty Ethics Assessment Committee of Humanities (FETC-GW); study approval registered under Bogae005-02-2017

the kick-off meeting, teachers and researchers exchanged ideas about the problems of the current reading comprehension curriculum (e.g., a lack of integration of reading instruction and other subjects, no attention to structures; see *Chapter 3*). Although the researchers took a primary role in defining the research problem, the teachers easily related to this problem and seemed eager to act as co-designers. In order to prepare teachers to their task, they were introduced to the four informational text structures, the design principles, and their legitimacy. In preparation of the second meeting, teachers were asked to select or write a suitable text and to develop a try-out lesson. During the second meeting, the teachers reflected on these try-out lessons, several design principles were clarified.

The teacher meetings were planned in such a way that teachers and researchers could first discuss practical issues, and then could focus on theoretical ideas and viability of the DPs. We expected teachers to gradually deepen their understanding of text structures, and become more aware of effective practices. Also, we expected them to adapt their beliefs and less effective habits in the ongoing process of adaptation and reflection during the design cycles (e.g., Bannan-Ritland, 2008; Hultén & Björkholm, 2016), especially because many DPs were different from current practice.

During the first design cycle, each teacher developed and tested five text structure lessons. Teachers reflected on each lesson in a logbook and received feedback on their artefacts from the principal investigator. Based on this first trial and feedback, teachers slightly adapted their lessons. Then, the researcher collected all lessons and finalized them for the second phase.

During the second design cycle, the teachers first exchanged experiences with the lesson development (i.e., text selection, choice of lesson goals, instructional approach) and discussed their ideas for refining the content and pedagogy of the materials. Then, they exchanged their revised lessons, so that they could teach five revised lessons of another teacher of the TDT. They kept a logbook and were observed during two lessons. During the fourth meeting, the teachers reflected on the main design issues.

Data collection

In order to analyze how teachers implemented the DPs, various types of qualitative data were collected (Table 2).

TABLE 2

Overview of lesson development, specifying the different roles and data sources

Design Cycle 1		Design Cycle 2		
Meeting 1 & 2	Try-out 1	Meeting 3	Try-out 2	Meeting 4
<i>Teachers</i>	<i>Teachers</i>	<i>Teachers</i>	<i>Teachers</i>	<i>Teachers</i>
1: Receive information on text structure and DPs.	Develop, test and revise first design.	Reflect on first design, discuss DPs and exchange lessons.	Test, revise and reflect on second design.	Reflect on second design and discuss DPs.
2: Apply DPs and discuss try-out design.				
<i>Researchers</i>	<i>Researchers</i>	<i>Researchers</i>	<i>Researchers</i>	<i>Researchers</i>
Discuss legitimacy of DPs, provide feedback on try-out design.	Provide feedback for revisions of first design. Support and finalize second design.	Moderate discussion and elaborate DPs.	Lesson observations.	Moderate discussion.
	<i>Data</i>	<i>Data</i>	<i>Data</i>	<i>Data</i>
	Artefacts and feedback, logbooks.	Panel interview.	Artefacts and feedback, logbooks, observations.	Panel interview.

Teachers were asked to comment in their logbooks on the development and implementation of each lesson. They reported on the time spent on lesson development and the duration of the lesson, and rated on a five-point Likert scale

their satisfaction with texts, pacing of activities, content, student and teacher activities. In addition, they had to mention their planned revisions for a second design. The logbook for the second design cycle contained additional questions about the completeness and clarity of the manual. During the second design cycle, the principal investigator observed and videotaped two lessons per teacher. Notes were taken about the duration of lesson phases, the implementation of instructional activities such as modeling and collaborative practice, and student involvement per lesson phase.

Teachers shared the first draft and the revised version of the lessons with the researchers. These artefacts gave insight in the series of adaptations the teachers made during the DBR. These adaptations and other issues relating to the design process were discussed during the third and fourth teacher meeting, in which two semi-structured panel interviews were held. The questions addressed the (1) text selection procedure, (2) choice of lesson goals, (3) satisfaction with researcher feedback, (4) instructional choices, and (5) perceived usefulness of the DPs.

In order to examine the viability of the design principles (RQ1), the researchers analyzed all artefacts and paid special attention to the ways in which teachers operationalized the DPs, making note of missing elements and marking everything that deviated from the DPs. Additional information in relation to RQ1 was obtained from logbooks, which revealed teachers' views on the viability of the DPs. During the panel interviews teachers were asked to elaborate on their comments.

The question about the support teachers need (RQ2) was mainly answered on the basis of data from the teacher meetings and panel interviews. The transcribed interview scripts were reread in order to make a list of teachers' uncertainties and questions relative to the design. The support question was also answered on the basis of a continuous process of reflection; the researchers reflected on the teachers' progress by comparing the quality of the first and second design, taking into account the support and extra explanations that had to be provided.

4.4 Results

The first part of this section discusses the viability of the four DPs in practice. The second part of this section describes what difficulties the teachers experienced while implementing the DPs, and how the researchers provided support.

Viability of design principles

Overall, teachers were satisfied with both designs. During the first design cycle, the teachers developed 20 lessons about four informational text structures. In the second cycle, these lessons were adapted and tested again, with major text revisions and, to a lesser extent, revisions in the content of the instruction.

DP1 required teachers to embed reading comprehension instruction in content-area subjects. This principle was implemented as a guideline for text selection: teachers chose texts close to the content-area curriculum. In the first design, seven texts were unrelated to content-area subjects, but were selected because of a clear text structure (e.g., a recipe). In the second design, most of these texts were replaced. DP2 required teachers to focus their instruction and use of reading strategies on text structure. The teachers were satisfied with DP2 and considered text structure a helpful tool for students to get the gist of a text and integrate information (1).

1) We always try to teach them that they must be aware of how the paragraph fits within the whole text, and how that fits with the title. By looking at the text from the perspective of text structure, I think my students learned more easily how to summarize at the text level and go beyond paragraphs. (Interview, T1)

Two teachers reported that, to their surprise, their students spontaneously applied their knowledge in other lessons. All teachers experienced a tension between DP1 and DP2: because they had difficulties finding and selecting well-structured content-area texts, they ended up with many poorly structured texts, which complicated text structure instruction. As texts were fundamentally revised in design 2, the number of texts with poorly structured texts were reduced.

DP3 required teachers to formulate declarative, procedural, and conditional lesson goals focused on reading. However, in 55 % of the lessons of the first

design, conditional goals were lacking. Instead, teachers often formulated content-related lesson goals focused on the text topic, not on reading outcomes. After feedback from the researchers, the lesson goals in the second design were refined, or eliminated if they were not focused on reading comprehension. When reflecting on DP3, the teachers mentioned that formulating lesson goals gave them more ownership and made them more reflective about their lessons (2).

2) When designing the lessons you really start to focus on those lesson goals. It makes you think: why would I actually do this? Because with those text books, well, you just simply recite the lessons. (Interview,T3)

DP4 required teachers to ensure a gradual release of responsibility (GRR) in their lessons, by following a sequence of explicit instruction, modeling and reflection, guided practice, collaborative learning, and individual practice. Table 3 summarizes how often each lesson phase was realized in each design, and its allocated time.

TABLE 3
Instructional approach: occurrence (%) and scheduled time (min) per component

Component	Design Cycle 1		Design Cycle 2	
	Occurrence	Time (SD)	Occurrence	Time (SD)
Explicit instruction	80	4.4 (2.9)	90	6.6 (4.8)
Modeling	90	5.1 (2.8)	90	5.3 (2.5)
Reflection	60	1.6 (2.0)	85	3.8 (2.4)
Guided practice	100	9.3 (3.9)	100	10.0 (4.7)
Collaborative learning	95	13.8 (8.6)	95	11.3 (8.0)
Individual practice	65	4.1 (3.5)	90	6.6 (4.1)

In the first design, teachers emphasized collaborative learning, guided practice, and modeling. Although the researchers had strongly recommended an explicit moment of reflection after modeling, this was realized in only 60% of the lessons in the first design. All teachers were eager to incorporate collaborative activities, in particular the jigsaw activities, and came up with many more types of student activities (e.g., games) than the researchers had suggested. By contrast, in 35 %

of the lessons, they did not incorporate individual activities. In the second design, more time was allocated to explicit instruction, and more lessons contained a reflection after modeling. Complex collaborative activities were replaced, and extra individual tasks were included, so that most revised lessons displayed a full GRR.

Implementation challenges and teacher support

Table 4 summarizes the issues that were encountered during the implementation of the DPs, and the data sources in which these were found. In the following subsections, we will explore these difficulties into more detail, and discuss the ways in which the teachers and researchers resolved these issues.

Unclear structure of content-area texts (DP1)

During the first design cycle, but also in the third teacher meeting, all teachers complained that it was almost impossible to find appropriate texts in content-area text books, as these texts often lacked a clear text structure (3). As a result, text selection became a time-consuming task that sometimes demotivated the teachers, especially when they had to revise the text in order to be able to use it for text structure instruction (4).

3) Soon I discovered that most texts did not have a clear, for students recognizable, structure. (Interview, T2)

4) I even developed lessons during my days off, because selecting texts was very time consuming. (Interview, T3)

After the first try-out, the teachers found out that selecting a text with a clear structure, or revising a text to the same end, was crucial for success (5). As text revision was difficult and time consuming, two teachers asked the researchers to take over text revision for the second design. The other teachers kept using authentic texts they hardly revised, although this negatively affected their text structure instruction (6).

5) Once the text is right, the lesson will be right. Then it suddenly becomes clear what you should teach in a lesson. (Interview, T4)

6) The text was so complex that we were not able to make and discuss the graphic organizer. (Logbook, T2)

Competing selection criteria (DP1)

The teachers experienced a tension between DP1 and DP2 in their text selection criteria. On the one hand, they wanted to select authentic content-area texts (DP1), on the other hand, they had to look for well-structured texts (DP2). The teachers had other selection criteria as well, such as appropriate sentence length and interesting topics (especially for boys). Although a clear structure was expected to be the main criterion, teachers gave priority to the text topic. However, after discussing the first lesson try-out in the TDT, the teachers became more aware of the fact that an adequate structure was a non-negligible, critical selection criterion.

Two teachers prioritized DP1: they selected authentic texts and made few revisions, as they were used to do at their schools. They simply “looked which text structure could fit best.” However, fitting in a structure was challenging, and simply adding signaling words did not fundamentally improve the text structure. Even in the second design, they kept using authentic, hardly revised texts, even at the expense of text structure instruction, as the lesson observations revealed: they spent much instruction time on explaining text content and vocabulary, so that little time was left for text structure instruction. They also struggled with exercises to represent the text with graphic organizers. During the second design cycle, they gradually became more critical about their own beliefs (7 and 8).

7) I had to explain a lot because of the conceptual difficulty of the text. Maybe the text is too complex to talk about text structure as well.
(Logbook, T1)

8) Using a Venn diagram to summarize a complex, authentic text felt like flogging a dead horse. I think that a more stereotypically structured text might have been a better choice. (Interview, T2)

Difficult to revise text structure (DP2)

As many content-area texts were not well structured, text revisions were needed. The researchers asked the teachers during the first design cycle to critically examine the texts, and to revise the top-level structure in order to make it more salient. The teachers did not realize this at first: only once, a teacher proposed that the text should be drastically edited (9).

9) I think the text should be rewritten. But it is too time consuming and complicated to change the top-level structure. (Logbook, T3)

Teachers indicated in their logbooks and in the panel interviews that text revision formed a major obstacle: they could not reorder ideas in the text in a way that would clarify the text's top-level structure. None of the teachers clarified the top-level structure by reordering information beyond the sentence level. Instead, they made local revisions, by adding signaling words. As a result, the majority of texts lacked a clear top-level structure or had a mixed structure (i.e., different structures within paragraphs without a clear overarching text structure).

It appeared that teachers' difficulties with text revision were due to their difficulties with recognizing text structure in the first place. Because of their strong focus on local text structure markers (i.e., signaling words), they did not always successfully identify the underlying top-level structure. For instance, a teacher treated a text with many dates as a chronological structure, whereas it actually focused on causes and effects of European collaboration. Because of teachers' difficulties, the researchers themselves revised the texts for the second design by reordering information at the paragraph level, and by simplifying complex vocabulary. These changes were discussed with the teachers during the third meeting. The logbooks show that teachers felt happier with the revised texts.

Reading strategies not tailored to structure (DP2)

During the first design cycle, text structure instruction seemed to be treated as an addition to the business-as-usual curriculum. Teachers kept teaching global reading strategies (e.g., making predictions based on pictures, instead of based on expectations about the structure), and to this, they added text structure

instruction. The lesson artefacts showed a lack of integration: teachers were not aware of the fact that they could demonstrate a structure-specific use of reading strategies such as predicting and questioning. Only with regard to summarization, teachers had no difficulties integrating this reading strategy with text structure: they provided structure-specific graphic organizers to help their students summarize the texts.

As teachers hardly integrated strategy instruction with text structure instruction, there seemed to be an overkill of information in the lessons. The teachers also overemphasized the presence of signaling words, possibly because signaling words formed a familiar topic for the teachers, and can easily be identified at the sentence level, without focusing on the top-level structure that was often rather unclear in the texts of the first design.

The teachers were not satisfied with this approach themselves: they were afraid it was monotonous, and could become a simple trick (10).

10) You must be careful; the children might simply underline signaling words, and then it becomes a trick they apply without thinking. You don't want that to happen. (Logbook, T2)

It appeared that DP2 should clarify that a focus on text structure also implies a structure-specific strategy use, in order to prevent too many angles on the text. Therefore, during the third meeting, the researchers elaborated DP2 in order to clarify that a focus on text structure meant that all strategy use should be tailored to the text structure at hand. The researchers explained again a structure-specific strategy use with concrete examples, and encouraged teachers to incorporate a larger variety of reading strategies – and not overemphasize signaling words. The researchers illustrated with revised modeling scripts how teachers could make better predictions about the content of the next paragraph based on text structure, and the kind of questions they could ask. The clarification of DP2 and additional support with concrete examples led to more integration in the second design cycle: teachers more often made predictions about the form and content of the text based on text structure, and varied more in their strategy use.

TABLE 4

Issues related to the implementation of each design principle

Design principle	Issues	Data source				Quote/Example
		A	L	P	O	
DP1 <i>Embed reading instruction in content-area subjects</i>	Unclear structure of content-area texts	✓	✓	✓	✓	<i>I searched content-area books first, but soon I discovered it would be time consuming to work with those texts. They were not suitable at all: texts were too simple, or it was narrative-like. Summarizing this text with a Venn diagram was like flogging a dead horse.</i>
	Competing selection criteria	✓	✓	✓	✓	<i>The texts had to match students' interests and level, and also contain enough leads to teach structure. That was challenging. I had to explain a lot because of the text difficulty. I think the text might have been too complex to also discuss structures.</i>
DP2 <i>Focus on informative text structures</i>	Difficult to revise text structure	✓		✓		<i>When I had a chronological text, I looked which signaling words the text already contained, (...) and I added extra signaling words.</i>

Design principle	Issues	Data source				Quote/Example
		A	L	P	O	
DP2 (continued)	Reading strategies not tailored to structure	✓		✓		<i>Look at the picture. What possible causes of drought can you think of? (No structure-based prediction)</i>
DP3	Content-related goals	✓		✓		Teachers formulate lesson goals focused on content.
<i>Balance declarative, procedural and conditional knowledge</i>	Limited attention for conditional goals	✓		✓		Difficulties with formulating conditional knowledge goals, and no alignment in lesson content.
DP4	Modeling: no coping model or student modeling	✓		✓		They only provide scripts for teacher modeling (mastery model, no coping model).
<i>Ensure a gradual release of responsibility</i>	Collaborative activities: individual accountability and poor transferability	✓	✓	✓		<i>I did not do the game with my class, because I actually didn't understand it myself.</i>
	Few opportunities for individual practice	✓		✓		Teachers ask students to work together on individual tasks.

Note: Data source is checked if evidence for implementation issue is found in A: artefacts, L: logbooks, P: panel interview, O: observations.

Foregrounding of content-related goals (DP3)

Teachers were tempted to emphasize content-related goals: for two teachers, understanding the meaning of the whole text was more important than acquiring new knowledge about text structure, as shown in (11).

11) After this lesson, students will know how the lives of Sherpa's changed over the past decades. (Lesson artefact, T3)

Even if reading goals were formulated, the teachers did not always match their instruction to it; they still emphasized the content of the text, and much less the text structure, even after researchers' feedback. Therefore, during the third meeting, it was stressed why it was important to focus more on reading goals and less on content-related goals. This had little effect, as the teachers kept emphasizing the content of the text in the second cycle, which in some lessons led to a limited alignment between the lesson goals and the actual lesson focus.

Poor operationalization of conditional knowledge (DP3)

The artefacts show that the teachers had difficulties formulating conditional knowledge goals (*when and why* lesson goal). Lessons lacked activities that could help students plan and evaluate their reading approach, and instruction on conditional knowledge consisted of simply telling students when and why a strategy would be useful. The concept of conditional knowledge appeared too unfamiliar for teachers to apply. During the third meeting, the researchers provided concrete examples of activities to promote conditional knowledge, for example by having students reflect on a reading approach in specific scenarios (12). Teachers gladly incorporated these tasks into their revised lessons.

12) Pete wants to summarize a text about the differences between viruses and bacteria. What kind of summary would you recommend, and why?

Only mastery models (DP4)

During the kick-off meetings, teachers and researchers discussed guidelines for effective modeling. Teachers were encouraged to sometimes act as a coping model. However, in the first lesson design, all teachers acted as mastery models.

As the researchers thought this might be due to a lack of knowledge, they encouraged teachers during the third meeting to examine the modeling scripts to see whether they could make some intentional mistakes and repair them while thinking aloud, in order to show their students why certain strategies do or do not work well in specific situations. This did not result in important changes in the second lesson design. Therefore, after the fourth meeting, in collaboration with one teacher, six video clips were produced to demonstrate the reading processes of weaker and stronger readers.

Issues with collaborative activities (DP4)

Teachers gladly followed the suggestion to design jigsaw activities as collaborative learning tasks. However, one issue became apparent during lesson observations: as the final exercise often consisted of students simply exchanging information without an additional task, their peers did not always feel the need to listen carefully. Therefore, after the fourth meeting, the researchers made sure that the final phase of jigsaw activities had a solid individual component (e.g., quiz or individual writing task) for which they really needed information from their peers.

Another issue concerned the transferability of the collaborative activities: the more creative the activity, the more difficult other teachers found it to carry them out in their own class. For instance, one teacher invented a game on the effects of import tax on trade. After the try-out, she described the success of the game. However, the teacher who tested it in the second design cycle, reported that she skipped the game because she did not understand it herself. Therefore, after the fourth meeting, two complex activities were changed into more familiar tasks.

Limited individual practice (DP4)

Although the researchers emphasized that collaborative activities should be followed by individual activities, the teachers left hardly any room for individual practice. During the third and fourth meeting, the interpretation of DP4 was discussed; teachers wondered if all phases of the GRR model should be realized in each lesson, or if these should be distributed over various lessons. Lesson observations also revealed that even if there was time for individual activities, teachers still assigned it as a group task. The issue was partially resolved by

strengthening the individual component of the jigsaw activities and adding individual activities aimed at conditional knowledge.

4.5 Conclusion and Discussion

Viability of design principles

Our design-based research provided insight into the viability of the design principles. Overall, the co-designing teachers seemed satisfied with the DPs, and reported to be more conscious about what they actually wanted their students to learn. The teachers experienced ownership over their lessons, as has also been reported in previous DBR studies (Cviko et al., 2014; Wikeley et al., 2005). Especially DP3 that required teachers to formulate lesson goals was highly appreciated.

It turned out that not all DPs were viable in their current form, and needed some elaboration. In DP2, the ‘focus on informative text structures’ did not state clear enough that it required teachers to demonstrate a structure-specific reading strategy use. In order to increase the viability of DP2, it seems wise to both emphasize this, and to provide more guidance in which strategies should be taught to guarantee a sufficient level of variation in strategy use. DP4 raised the question whether all lesson phases from the GRR model should be part of each lesson, or if and how they should be distributed over multiple lessons. This raises an important topic: although the GRR can support a structured design of instruction and activities, it is also important that there is room for teachers’ professional flexibility to fit instruction to students’ needs and practical demands (Webb et al., 2019).

In addition, the simultaneous implementation of DP1 and DP3 proved difficult. Although teachers were encouraged to embed reading in the context of the content-area classroom (DP1), they also had to primarily focus their instruction on reading-related lesson goals, and not content-related goals (DP3). The effects of this complex integration were amplified by the fact that teachers themselves were unaware of this tension, and did not reflect on it until the researchers signaled the issue. With help from the researchers, the integration of both design principles and the alignment of lesson goals and activities gradually improved, and the focus shifted from content goals to reading-related goals. However, the way in which linguistic and non-linguistic learning goals can be combined is not

easy to determine (Nikula et al., 2013; Vázquez, 2014), although text structure instruction might provide a good middle ground to do justice to both specific knowledge-building lesson goals and generalizable reading-related goals (Read et al., 2008; Reutzel et al., 2005; Williams et al., 2014).

The lesson design was complicated by the fact that teachers hardly found well-structured texts in their content-area books. This might raise questions about the ecological validity of the lesson series: if students rarely encounter well-structured texts in their books, why bother about teaching them about text structure? First, the experienced lack of suitable materials might have partially been due to teachers' limited experience in recognizing structures. Second, the low number of clearly structured texts might reflect the limited quality of Dutch content-area textbooks which often lack coherence markers such as *because* or *that's why* (Land et al., 2007). Perhaps educational publishers need to become aware of the need to provide well-structured texts. Third, and most importantly, even if current textbooks display relatively few basic structures, it remains important to familiarize students with the basic structures as described by Meyer (1975). They are foundational for students to understand more complex and combined structures, and form an important stepping stone to successful reading of authentic, less structured texts. It is important to provide an optimal sequence that matches students' zone of proximal development, as empirical research suggests that some structures are easier to learn than others (Meyer et al., 2018).

Support for co-designing teachers

It is often advocated to aim for a collaborative partnership on equal terms between teachers and researchers in a DBR project (Bednarz et al., 2012; Broekkamp et al., 2007), as it results in a democratic type of epistemology with co-constructed knowledge. The theory-driven knowledge, views, and experiences of the researcher(s) are mixed with the context-based experiences, knowledge, and routines of practitioners in a continuous process of negotiation and reflection (Bednarz et al., 2012). This might promote teacher professionalization (Broekkamp & van Hout-Wolters, 2007; Hultén & Björkholm, 2016; Kafyulilo et al., 2016).

In our DBR, there was indeed a valuable complementarity between teachers and researchers. Researchers took the lead by providing a set of design principles that functioned as a stepping stone between research and practice, whereas

teachers had the liberty to experiment with the design principles quite independently during the first design. This collaboration was very valuable, as it led to an interesting mix of practical and theoretical insights in the final prototype. Although the researchers were mainly focused on evidence-based lessons, the teachers felt the need to make sure that students would feel engaged. Therefore, they selected challenging text topics, and came up with many more engaging activities than the researchers could have imagined. Thanks to teachers' creativity, students appeared motivated and engaged.

All teachers were well aware of the problems in the context of reading comprehension, but due to the existing research-practice gap, their knowledge about evidence-based reading comprehension instruction was limited. As a result, there was some asymmetry in power between teachers and researchers in this relatively short DBR project. The collaborative relationship between teachers and researchers displayed characteristics of a tutor-tutee relationship, which was slightly different than in the original sense of the intended reflexive contract (Bednarz et al. 2012), but this is not uncommon in small-scale DBR projects that involve primary school teachers (Ormel et al., 2012). Over time, the teachers in the TDT gained more knowledge: they benefited from continuous feedback, discussions during meetings, and most importantly, from DPs with very concrete examples. In long-term collaborations, this phase of intensive support may be followed by a phase in which a more equal partnership develops over time.

The challenging nature of co-design became apparent in the high amount of guidance and feedback teachers needed to successfully implement the DPs, most notably with regard to text selection and revision. Teachers were faced with poorly structured content-area texts, and/or held conflicting text selection criteria in which they esteemed an engaging topic more important than a suitable text structure. Teachers felt unable to clarify the text structure beyond the sentence level. Another issue that was at least partially caused by teachers' limited knowledge was a poor alignment between text, lesson goals and activities, and to only a superficial integration of the design principles in the first design. That is, teachers emphasized content-related lesson goals at the expense of reading-related goals, had no idea what kind of activities could promote conditional knowledge, and often did not tailor their reading strategies to the structure at

hand. This issue was only resolved through intensive feedback, many examples, and discussions with the researchers.

Asking teachers to provide well-structured texts themselves, designing lesson goals and activities while obeying various pedagogical guidelines as well, seemed too demanding for the teachers. The teachers needed more pedagogical content knowledge: a specific kind of knowledge that is neither pedagogy nor content per se, but combines both in a unique way (Gudmundsdottir & Shulman, 1987). That is, the teachers in our project were able to explain and demonstrate reading strategies as they were used to, but when this had to be combined with specific knowledge about text structure, it soon became too challenging. This is, however, no real surprise: previous research has already shown that teachers struggle to recognize and teach text structures (Beerwinkle et al., 2018; Reutzel et al., 2016). Furthermore, the Dutch teacher training curriculum is very implicit about text structures (Kooiker-den Boer et al., 2019), and teachers do not usually encounter good examples in current textbooks for reading comprehension instruction (*Chapter 3*).

Other factors influencing the design process

It appeared that the design process was also affected by teachers' beliefs and habits. Researchers and practitioners might hold different values and beliefs (McKenney & Reeves, 2018; Voogt et al., 2011), for instance about the desired content or instructional approach, or about their role as co-designer (Cviko et al., 2014). In our project, teachers' beliefs and habits influenced both the content and the pedagogy of the designed lessons. For instance, teachers emphasized collaborative activities at the expense of individual practice. Possibly, teachers were simply very enthusiastic about collaboration, as individual practice is very characteristic of current practice, and collaboration might have appeared very innovative (*Chapter 3*). Also, the teachers were reluctant to act as coping models: instead, they always acted as mastery models as they were used to do, possibly because they were afraid to set a wrong example.

Teachers' instruction was also influenced by beliefs and habits: two teachers believed that authentic texts and intensive vocabulary instruction were of utmost importance. This interfered with text structure instruction, for which well-structured, not too complex texts were needed, so that students could solely focus

on structure, without being distracted by difficult concepts. Also, all teachers were tempted to prioritize content-related lesson goals over reading goals, and actually teach geography, instead of reading comprehension.

With a shared effort, a DBR can generate insight into design principles, stimulate teacher professionalization, and lead to an improvement of curricular products, thereby contributing to bridging the research-practice gap. However, it is a challenge to engage teachers as co-designers, because they are often not aware of their limited knowledge or different perspectives. For a successful project, both teachers and researchers have to invest. From teachers, it requires an open mindset to learn; from researchers, it requires them to not only assume a facilitative role, but also provide adequate support and clear design principles that form a concrete stepping stone between theoretical insights and classroom practice. Also, researchers should be sensitive to teachers' habits and beliefs, and make an effort to discuss the value of these beliefs in relation to the design project at hand, as they might otherwise interfere with the implementation of design principles.

More research should find its way into the classroom. Still, it is questionable whether the wide scope of knowledge and skills required for successful co-designing can actually be expected from teachers. It is recommendable to engage educational publishers to provide high-quality content-area texts (see also Chambliss & Calfee, 1998; Dewitz & Jones, 2013), before teachers start designing lessons. This might turn the content-area classroom into a fruitful context to teach reading comprehension. Co-designing might not happen in collaboration on equal terms, but with a long-term, shared effort from teachers and researchers, it can contribute to teacher professionalization, and to improved curricular materials. After all, Rome wasn't built in one day – and so are the bridges we try to build between research and practice.

CHAPTER 5



Text structure instruction in primary education: Effects on reading, summarization, writing, and metacognitive knowledge

Abstract

Although knowledge of informational text structures can promote deeper text comprehension, this topic receives little attention in the Dutch primary school curriculum. Dutch students ($n = 201$) in grades 4-6 participated with their teachers ($n = 10$) in this quasi-experimental study with a switching-panels replication design. Students either first followed a text structure intervention (TOP) and then went back to business-as-usual, or the other way around. During the intervention, teachers taught their students about the characteristics of four informational text structures, and how to use structure-specific graphic organizers to organize main ideas for each structure. In addition, several writing tasks related to the different text structures were included. Only the fourth graders in one iteration of the intervention showed immediate effects over and above the effect of business-as-usual lessons on the text structure test ($d = 0.50$), the reading comprehension test ($d = 0.53$), the summarization task ($d = 0.48$), and the explicit metacognitive knowledge test ($d = 0.24$). In both iterations of the intervention, an immediate effect was found on writing ($d = 0.33$ and $d = 0.39$). These findings are discussed in the light of test-related issues and implementation fidelity data.

5.1 Introduction

Despite intensive reading instruction, a substantial number of students struggle with reading comprehension, especially when it comes to deeper processing of informational texts (Diakidoy et al., 2003; Helder et al., 2016; Kim & van Dusen, 1998; Kraal et al., 2018; Rooijackers et al., 2020; van den Broek et al., 2021; Wijekumar et al., 2020). This problem also holds true for the Dutch context where many students appear to struggle with integrating information (PIRLS-2016: Gubbels et al., 2017; PPON-54: Kuhlemeier et al., 2014), as well as with summarizing texts (PEIL.Leesvaardigheid-2020/2021: Dutch Inspectorate of Education, 2022). These comprehension problems often become apparent on the threshold of the upper elementary grades when the focus shifts from learning-to-read to reading-to-learn (Chall & Jacobs, 1983; Harlaar et al., 2017).

One reason for these comprehension problems may be the sudden increase of informational texts to be read around fourth grade. Informational texts are generally assumed to be less familiar and more complex than narrative texts. Besides a difficult vocabulary and high information density, they also have complex text structures (Coté et al., 1998; Wu et al., 2020), such as compare-contrast, chronology, problem-solution or cause-effect (Table 1; see also Meyer 1975). Students are often less familiar with such structures than with the structure of narrative texts (Graesser et al., 1994; Hiebert & Mesmer, 2013; Read et al., 2008).

A second reason might be that reading-to-learn requires deeper text comprehension skills: students need to make connections across different parts of text, identify main ideas and relate these to their prior knowledge, and evaluate what they read (Biancarosa & Snow, 2004; Harlaar et al., 2007; Moort et al., 2020). This echoes the Construction-Integration Model (Kintsch, 1988, 2004, 2013; van Dijk & Kintsch, 1983), which states that, for deep comprehension, readers parse textual input into concepts and relationships, which they need to organize in associative networks. Ideally, readers make connections between ideas within the text, and integrate these ideas with prior knowledge, until a coherent mental representation of the text arises: the so-called situation model.

Situation model comprehension requires that readers carefully think about the main ideas and how these are linked across the text. This process can be fostered by insight in the underlying text structure (Kendeou & van den Broek, 2007; Meyer & Freedle, 1984; Wijekumar et al., 2020). That is, a reader who

recognizes the underlying text structure is like a traveler with a road map that highlights the main routes and gives insight into what is coming next. Likewise, knowledge about text structures helps readers predict upcoming information, and identify and organize main ideas and their implicit and explicit relationships more easily (e.g., Meyer & Ray, 2011; Meyer et al., 2002; 2011; 2018; Strong, 2020; Wijekumar et al., 2012; 2013; 2014; 2017; 2019; Williams et al., 2004; 2014; 2018). For instance, students who read *due to* in a science text should infer that a cause-effect will follow, and students who recognize a problem-solution structure will realize that after the presentation of the problem, the text will probably discuss several solutions, even if these are not explicitly highlighted with signaling words.

TABLE 1

Overview of the four text structures with signaling words and phrases in italics.

Texts with compare-contrast structure explain how things are similar or different. Influenza <i>and</i> COVID-19 are <i>both</i> contagious respiratory illnesses, <i>but</i> they are caused by <i>different</i> viruses. COVID-19 is caused by a coronavirus, <i>whereas</i> flu is caused by an influenza virus. The symptoms of flu and COVID-19 are <i>similar</i> : <i>both</i> cause symptoms like fever and shortness of breath.	Texts with problem-solution structure explain a problem, and how it can be fixed. The COVID-19 pandemic resulted in a worldwide shortage of face masks. Also, many people were furloughed and needed jobs. A social enterprise <i>addressed both of these problems</i> : train workers in India started producing face masks. <i>This way</i> , people had jobs and masks were produced.
Texts with cause-effect structure tell how an event culminates in an outcome or effect. Scientists still aren't sure why COVID-19 is <i>causing</i> clots. Clots might be <i>the result of</i> blood vessels' <i>reactions to</i> being invaded by the virus. The virus attacks cells via a receptor called ACE2. <i>When</i> the virus binds to these receptors in the walls of blood vessels, they <i>become</i> inflamed, which <i>can cause</i> clotting.	Texts with chronological structure explain an order or cycle of events and/or procedures. <i>On 1st January 2020</i> , the Wuhan seafood market was closed for inspection, as it appeared that many visitors developed a new disease. <i>On January 7</i> , Chinese authorities identified that the disease was caused by a new coronavirus. <i>Five days later</i> the Chinese government shared the genetic sequence of the virus.

As knowledge about text structure can foster readers' higher-order text comprehension, it seems promising to explicitly teach students about these text structures. Various researchers, mainly in the US, have developed intervention programs in which students learn to recognize text structures, use structure-specific main idea sentences or graphic organizers to summarize text, and make structure-specific inferences about text content (e.g., Meyer et al., 2002; Wijekumar et al., 2012; Williams et al., 2004), sometimes followed by structure-specific writing tasks (e.g., Strong, 2020). In fact, these interventions have intertwined explicit knowledge about text structures with reading strategy instruction, a well-researched approach to support comprehension (Cromley & Azevedo, 2007; Palinscar & Brown, 1984; Soodla et al., 2016). Such an approach might promote a contextualized use of reading strategies, that is, a strategy use that is aligned with the specific text structure at hand (Hoch & McNally, 2020; Stevens & Vaughn, 2021).

Although this approach of providing text structure instruction appears effective for students of various ages (*Chapter 2*; Hebert et al., 2016; Pyle et al., 2017), and abilities (Gajria et al., 2007), several questions remain that require additional research. First, the generalizability of findings to other populations and educational contexts is relatively unclear (Bohaty et al., 2015). The number of intervention studies outside the US is gradually increasing (e.g., Chile: Ponce et al., 2012; China: Cheng, 2018; Xu et al., 2021; Indonesia: Eliata & Miftakh, 2021), but the majority of text structure research is still US based. It is important to study the effectiveness of text structure interventions in other linguistic and cultural contexts with varying reading curricula and teacher training programs (Bohaty et al., 2015; Williams, 2018; Wijekumar et al., 2021).

The current study takes place in the Netherlands where the primary school curriculum is very implicit about informational text structures (*Chapter 3*). As there are reasonable concerns on Dutch students' reading achievements, in particular with respect to higher-order text comprehension (e.g., PIRLS-2016; Gubbels et al., 2017; PISA-2018; Gubbels et al., 2019; van den Broek et al., 2021), this country provides an interesting context to study the effects of text structure instruction.

Second, a plea has been made that text structure research should include multiple outcome measures for reading, and examine transfer effects more carefully (Bohaty et al., 2015; Hebert et al., 2016; Williams, 2018). That is, the effects

of text structure instruction on text comprehension vary greatly per type of outcome measure (*Chapter 2*; Hebert et al., 2016; Pyle et al., 2017) and potential transfer effects on metacognitive knowledge and writing skills are much less examined. Yet, explicit text structure knowledge could foster students' metacognitive knowledge and strategic reading (Collins, 1994; Gordon, 1990; Zarrati et al., 2014) and has been shown to promote writing skills (Hebert et al., 2018; Kirkpatrick and Klein, 2009; Strong, 2020). Furthermore, studies that measure delayed effects beyond the week after finishing the intervention are scarce (Hebert et al., 2016). Therefore, our study includes a broad spectrum of reading tests, while exploring effects on writing and explicit metacognitive knowledge as well. The switching panels replication design (Shadish et al., 2002) of the current study allows us to replicate the intervention and test for maintenance effects 13 weeks after the intervention (see Method section).

Third, text structure research should examine powerful instructional approaches: now that we know *what* to teach in text structure interventions, we need to know *how* to teach about text structures most successfully (Hebert et al., 2016; Williams, 2018). According to Pyle and colleagues (2017), instruction that follows a gradual release of responsibility (GRR; see Fisher & Frey, 2013; Pearson & Gallagher, 1983) might have the biggest impact on reading outcomes, but their meta-analysis did not examine instructional features. Instructional features were examined in the meta-analysis in *Chapter 2*, which suggested that individual practice might affect the ultimate maintenance of effects, but also highlighted how the instructional approach is often poorly described and operationalized. For example, only a handful of text structure interventions use structured forms of collaborative learning, even though high-quality collaborative learning activities can facilitate scaffolding, promote effective student interactions, and increase student motivation (Fuchs et al., 1997; 2021; Garibaldi, 1979; Nolte & Singer, 1985; Palinscar & Brown, 1984; Puzio & Colby, 2013). Therefore, the current intervention followed a GRR pattern, in which ample opportunities for collaborative learning were provided. In sum, this chapter will focus on the following research question:

RQ: What are the effects of the text structure intervention program TOP on reading comprehension, text structure knowledge, summarization skills, explicit metacognitive knowledge, and writing for fourth and fifth graders?

5.2 Method

Intervention (TOP) and business-as-usual (BAU)

The text structure intervention program TOP (see Appendix C) was developed in close collaboration with four co-designing primary school teachers (see *Chapter 4*). The TOP intervention consists of ten one-hour lessons on four text structures, subsequently on the compare-contrast, problem-solution, chronology, and cause-and-effect structure.³ During the intervention, lessons were completed at a rate of one or two lessons per week, while the business-as-usual group continued their existing routines.

Table 2 summarizes the content and the instructional approach of the TOP intervention. The first lesson introduced students to the concept of text structure and emphasized why it is useful to recognize text structure. This was followed by two or three lessons that focused on each structure's main components and key characteristics, the structure-specific graphic organizers, and how students could use certain reading strategies (questioning, summarizing, predicting) in the context of that structure (e.g., *Which questions are useful to ask while reading a compare-contrast text?*). The texts were related to various content-area subjects: eight texts were related to the biology curriculum, five to geography, and three to history as this provides an effective and meaningful context for reading comprehension instruction (Hwang et al., 2022; Maerten-Rivera et al., 2016) and might promote transfer and motivation (Duke, 2020; Williams et al., 2014).

Each lesson consisted of five instructional phases that were ordered according to the Gradual Release of Responsibility instructional model (Fisher & Frey, 2021; Pearson & Gallagher, 1983; Webb et al., 2019). During explicit instruction, teachers briefly introduced or reviewed one text structure. Then, teachers followed a modeling script and demonstrated how the structure could be recognized (i.e., how to look for text-structural cues), and how reading strategies (e.g., predicting, summarizing) should be applied in that specific structure. In three lessons, videoclips were provided in which a stronger and a weaker model read the same text while thinking aloud. After this, students and

³ This study was approved by the Faculty Ethics Assessment Committee of Humanities (FETC-GW); study approval registered under Bogae005-02-2018







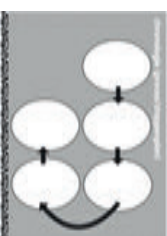
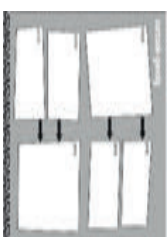
teachers engaged in a moment of reflection or abstraction (Schutz & Rainey, 2019), where they discussed their teachers' reading approach (e.g., *What strategies did I apply during reading, and why?*) or compared the video clips (e.g., *Which student had the best approach, and why?*). Such an active reflection after modeling explicitly focuses students' attention at crucial aspects of the reading approach, which seems to be related to improved student outcomes, at least in the area of writing (Braaksma et al., 2002; Rijlaarsdam et al., 2005).

Then, the role of the teacher was faded-out: after guided practice, students collaborated in groups to read the remainder of the text and work on a structure-based summarization or recall task. This collaborative activity was often designed as a jigsaw activity: students first worked in expert groups on one part of the text, and then formed mixed groups to exchange information with students who read another (part of the) text, to finally complete a joint product (Aronson, 1978). Intervention lessons ended with an individual activity, often a writing task (i.e., writing a paragraph in a specific structure) or a metacognitive reflection task (e.g., *Can you provide tips for creating a good Venn diagram? Can you think of other situations where you have to read chronological texts?*) in order to explicitly teach for transfer (Patton et al., 2022).

The control condition consisted of business-as-usual lessons (BAU). In these lessons, students followed their regular reading comprehension curriculum (e.g., *Nieuwsbegrip* [Understanding news], *Leeslink* [Reading link]). These highly-scripted curricular materials emphasize the practice of reading strategies related to different phases of reading (before, during, after reading) by providing texts with questions, but generally lack explicit instruction about informational text structures (Chapter 3). The observed BAU lessons focused, for instance, on procedural knowledge related to using a reading strategy such as dictionary use, or how to make predictions before reading a text. The main distinguishing feature of BAU versus TOP lessons was the lack of explicit attention to informational text structures.

TABLE 2

Intervention program TOP

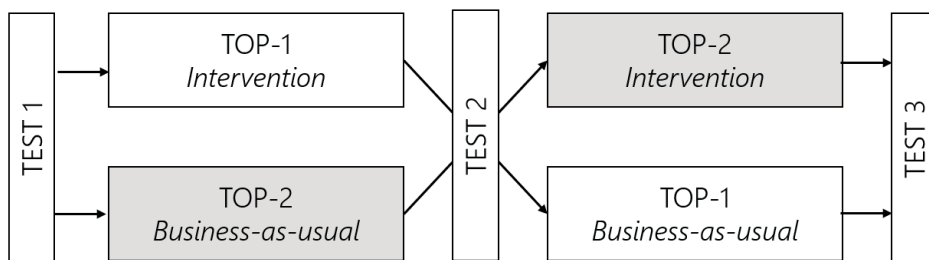
Introduction	Compare-contrast	Problem-solution	Chronology	Cause-effect
Lesson 1	Lesson 2 & 3	Lesson 4 & 5	Lesson 6, 7 & 8	Lesson 9 & 10
				
<u>Texts</u> Organ systems Seeds and spores Jaguars or cheetahs?	<u>Texts</u> Mustelids, rodents, or pack animals. Viruses or bacteria? North or South Pole? The Olympics: now and then.	<u>Texts</u> Historic inventions Plastic soup	<u>Texts</u> Recipe Sewerage Cocoa to chocolate Canals in Utrecht	<u>Texts</u> The Himalayas Impact of El Niño
<u>Activities</u> Sorting tasks	<u>Activities</u> Write summary Prepare a pitch	<u>Activities</u> Write summary Prepare a pitch Write problem-solution text	<u>Activities</u> Summary and quiz Write persuasive letter Write chronological text	<u>Activities</u> Write summary Play a class game Information poster
				

Research design

Intervention effects were examined with a switching-panels replication design (Shadish et al., 2002) with two groups and three measurement occasions (T1; T2; T3; Figure 1). Classes and teachers were randomly assigned to one of two groups. The TOP-1 group implemented the intervention (TOP) between T1 and T2 and followed business-as-usual lessons (BAU) between T2 and T3. The TOP-2 group started with their regular reading curriculum, and implemented the intervention program between T2 and T3. The research design has various advantages with regard to a single quasi-experimental design. First, as the research design enables us to implement the intervention in both groups, important information can be derived related to the replicability and internal validity of the intervention effects. Second, because of the three measurement occasions, maintenance effects can be assessed: T3 serves as a delayed posttest for the TOP-1 group. Third, the design is useful in educational research where random assignment to conditions might be unfeasible and/or unethical (e.g., Bouwer & Koster, 2016; Rogiers et al., 2021).

FIGURE 1

Switching-panels replication design with two groups: TOP-1 and TOP-2.



Sample

Ten upper-elementary school teachers (80% female) in seven Dutch primary schools volunteered to participate in the current study. They had on average 9.3 years of teaching experience ($SD = 5.9$) and taught in grade 4, grade 5, or in a mixed grade (5/6). Most schools were religiously affiliated (3 Catholic, 2 Protestant, 2 public schools), which is representative of the Dutch school system. The public schools participated with respectively two and three classes; the other schools with one class.

Overall, 203 students (55% female) participated in our study: 114 fourth graders (divided over four classes), 91 fifth graders, and 22 sixth graders (divided over three fifth grade classes and three mixed grade classes). The average class size was 20.7 ($SD = 6.1$). Classes were randomly assigned to either the TOP-1 group (starting the intervention at the beginning of the school year), or the TOP-2 group (starting the intervention after the Christmas break). Four classes somewhat overrepresented students with a lower socio-economic and migration background; these were equally distributed over both groups.

Table 3 provides various class-level and student-level features per group and grade. The groups were highly comparable, except for the fact that the teachers of the fourth graders in the TOP-2 group on average had less teaching experience than those of the fourth graders in the TOP-1 group ($p = .02$). Students with dyslexia ($n = 14$) were equally distributed over both groups ($p > .05$). No other information on special educational needs was provided.

TABLE 3
Characteristics per group and grade

	TOP-1		TOP-2	
	Grade 4	Grade 5+ **	Grade 4	Grade 5+
Students	48	60	40	53
Classes	2	3	2	3
Class size	21.0 (0.0)	21.0 (9.6)	20.0 (4.2)	21 (3.6)
Male students (%)	54	47	43	57
Teacher experience (years)	16.0 (5.7)	11.0 (6.1)	3.0 (2.8) *	7.3 (2.1)

Note. TOP-1: text structure lesson series before business-as-usual; TOP-2: vice versa.
* Means differed significantly between groups, with $p < .05$. ** Grade 5+ concerns 91 fifth graders and 22 sixth graders who participated due to mixed grade groupings.

Measures

Students completed five tests at three measurement occasions: a text structure test, a standardized reading test, a summarization task, a writing task, and a metacognitive knowledge questionnaire. Tests were administered one week prior to the start of the experiment (T1), and within two weeks after the end of the intervention (T2 and T3). As the estimated time for completing all tests was 2

hours, teachers spread the tests over three sessions within one week. Students could work without a time limit, but even so, some students failed to complete one or two tests at the first ($n = 12$), second ($n = 10$), or third ($n = 13$) measurement occasion.

Text structure test

The text structure tests that we designed (see Appendix C) mainly consisted of multiple-choice questions ($n = 10$ for T1, $n = 9$ for T2 and T3) that required students to recognize the underlying structure of text fragments or to make structure-based predictions based on a title, or to complete a main idea sentence with a short open-ended answer. In addition, for two content-area texts ($M_{length} = 207.5$ words, $SD = 33.0$), students received a list of numbered main ideas and a graphic organizer in which they had to put the numbers of the main ideas in the right boxes. The graphic organizer questions were inspired by sorting tasks that were used in previous studies to measure situation model comprehension (Land, 2009; Kamalski, et al., 2005). Students completed a problem-solution and a cause-effect chart at T1, a timeline and Venn diagram at T2, and a timeline and problem-solution chart at T3. Students received 1 point for each correct box. The maximum number of points to be earned per graphic organizer varied from 6 to 10.

For scaling purposes, the text structure tests were piloted among fourth and fifth graders in three non-participating schools ($n = 98$). The analyses showed that the test items had similar mean p -values, but that rit -values were rather weak. Therefore, some malfunctioning items were deleted (two items on T1, two on T2, and five on T3). The final reading tests were comparable in difficulty; p -values (.59, .59, and .58) and variances (.24 for all tests), but still not completely parallel in terms of their rit -values (.55, .30, and .59). Therefore, we concluded that the text structure tests could be used to reliably detect interaction effects resulting from the experimental manipulation, but that changes in scores between measurement occasions should not be interpreted as absolute growth in student abilities. Reliability analysis produced acceptable to good alpha-coefficients ($\alpha = .73$, $\alpha = .80$, and $\alpha = .86$).

Standardized reading test

The standardized reading tests contained expository passages and one narrative passage with multiple choice items and open-ended questions that pertained to literal text base comprehension or situation-model representation. The test items ($n = 20$ per occasion) were assembled from a database of yet unpublished PPON-items that are used for Dutch periodical nationwide reading assessments (see Kuhlemeier et al., 2014). All of these items had already been piloted on a large scale with third and sixth graders. As we wanted to interpret the changes in test scores between measurement occasions as actual growth in reading skills, we created three parallel reading tests by assembling items so that they were comparable in average p -values (.68, .68, and .65), variances (.22, .22, and .23), and rit -values (.33 at all three occasions). Reliability analyses produced more or less acceptable alpha-coefficients per measurement occasion ($\alpha = .59$, $\alpha = .65$, and $\alpha = .64$). We had to delete one item at T2, as analyses showed this item did not function properly.

Summary

At each measurement occasion, students summarized an informational text, consisting of an introduction to a problem, and three paragraphs with possible solutions ($M_{length} = 302.2$ words, $SD = 12.3$). The text topics related to different content-area subjects: geography (traffic jams) on T1, history (the Dutch water defense system) on T2, and biology (venomous animals) on T3. Dotted lines on the answer sheet provided implicit guidance about the desired summary length.

A scoring protocol was developed for each text by the first and second author. Summaries were first parsed into idea units (i.e., sentences that contained at least one verb), and each idea unit was then classified as main idea (1 point) or detail (0 points). For each text, students could score a predefined number of main ideas that was comparable across measurement occasions, and related to the crucial parts of a problem-solution text, such as *Problem*, *Solution-1*, *Solution-2*, and so on. The percentage of correctly included main ideas was used as outcome measure. Students did not receive points for redundant or false information. In order to check the reliability of this coding procedure, for each summary task, 32 summaries (16%) were also scored by one of two trained research-assistants, which revealed substantial agreement between raters (87%, 90%, and 81%).

Metacognitive knowledge

At all three measurement occasions, students' explicit metacognitive knowledge was assessed with a shortened version of the questionnaire originally developed by Brand-Gruwel (1995). This consisted of fifteen multiple choice questions about reading strategies that can be applied before, during or after reading (e.g., *What should you do when you don't understand a difficult word?*). Students had to circle the best option out of three: an adequate strategy, a less adequate strategy, and an irrelevant strategy (e.g., *Copying the answer from a friend*). Reliability analysis produced more or less acceptable alpha-coefficients ($\alpha = .61$, $\alpha = .59$ and $\alpha = .51$).

Writing

At each measurement occasion, students completed one writing task, taken from the Tekster intervention program (Bouwer & Koster, 2016) in which they had to write a persuasive letter (e.g., *Convince your teacher that you and your classmates need a classroom pet*). The three writing tasks were comparable with regard to the intended audience and the communicative goal. Two raters independently assessed the quality of each text by using a continuous rating scale (0-5) with five benchmark essays that represented the range of writing quality that can be expected of students in the upper elementary grades (Bouwer & Koster, 2016). The final text quality score was obtained by calculating the mean score of the two raters. Interrater reliability analysis showed agreement percentages of 80%, 86%, and 85%.

Treatment fidelity

As it is important to gain insight into teachers' enactment of reading interventions (Aarnoutse & Schellings, 2003; Beerwinkle et al., 2018; Bohaty et al., 2015; Okkinga et al., 2018b; 2021; Pyle et al., 2017; Turcotte et al., 2015; Wijekumar et al., 2019), we collected some data related to implementation fidelity and social validity.

With regard to the business-as-usual condition, all teachers ($n = 10$) completed a questionnaire on their regular curriculum. During the study, systematic business-as-usual lesson observations were carried out ($n = 10$) to carefully describe the content and instructional approach of the control condition, and to check that there was no explicit text structure intervention in this condition. For the group that followed the intervention first and then went back to business-

as-usual, this observation was relevant too, as it provided a check that teachers would not continue practices from the intervention program during business-as-usual lessons.

Following the same procedure, we also collected data on the implementation of the intervention by observing two intervention lessons per teacher ($n = 20$). We used an observation sheet (see Appendix E) to evaluate for each lesson phase (i.e., explicit instruction, modeling and reflection, guided practice, collaborative learning, individual practice, closure) its occurrence and duration (in minutes), and to check the quality of key components (1: implemented as intended; 0: not implemented as intended). For example, for collaborative learning, it was checked whether the teacher first created expert and then mixed groups (which is required for jigsaw tasks), and whether the teacher provided spontaneous feedback during the task. Due to ethics regulations, lessons could not be videotaped, but the validity of the observation scheme was increased by formulating concrete events that were evaluated on occurrence and duration, and by leaving room for a general impression and for remarks, so that difficult decisions could be discussed afterwards (Miles et al., 2013).

We also gathered data on social validity by collecting teacher entries in logbooks about each TOP lesson, asking them to evaluate the lesson, and to rate on a Likert scale their lesson appreciation, and the estimated level of difficulty of the lesson. At the end of the TOP program, we conducted retrospective interviews with the participating teachers. Due to unforeseen circumstances, we were only able to collect logbook ($n = 6$) and interview data ($n = 6$) of half of the teachers.

Analysis

Multilevel models were applied to the data. For the analysis of effects on the summarization task, explicit metacognitive knowledge test, and the writing task, linear multilevel models were applied on sum scores (i.e., the total score on each test per occasion) within a hierarchical three-level structure, with measurement occasions (T1, T2, T3; level 1) nested within students (level 2), who were nested within classes (level 3) (Hox et al., 2017). By stepwise model fitting (see Appendix D), we added as fixed effects Time (three measurement occasions: T1, T2, T3), Group (TOP-1 or TOP-2), Grade (4 or 5+) and their interactions to estimate means per group and measurement occasion, with variation within and between students

and classes as random effects. We examined the immediate effects of the intervention (T1-T2 for TOP-1; T2-T3 for TOP-2), and the delayed effects of the intervention as well (T2-T3 for TOP-1). In the case of interaction effects, custom hypotheses were applied to localize the effect (between T1-T2, and/or between T2-T3).

Generalized binominal multilevel models were applied to the binominal data of the text structure test and the standardized reading test. Because of the heterogeneous nature of the items, an additional level of items nested within measurement occasions was used, resulting in a hierarchical four-level structure (i.e., score on each single item per task per student per measurement occasion), with items (level 1) nested within measurement occasions (T1, T2, T3; level 2), nested within students (level 3), who were nested within classes (level 4). Following a similar approach to estimate means per group, we examined the immediate effects of the intervention (T1-T2 for TOP-1; T2-T3 for TOP-2), as well as the delayed effects of the intervention (T2-T3 for TOP-1). Variances for items and students within classes, and the variance between classes were modeled as random effects.

5.3 Results

Treatment fidelity

The content and pedagogy of the TOP and BAU lessons were evaluated by combining data from a questionnaire, lesson observations, logbooks, and interviews. On a 10-point scale, teachers gave a positive overall rating of the intervention program ($M = 8.14$, $SD = 1.03$). In particular, they appreciated the modeling scripts and the structure-specific graphic organizers, which students even applied spontaneously during content-area lessons, according to some teachers.

Table 4 summarizes the time that was allocated to different lesson phases, and indicates in what percentage of observed lessons certain instructional phases were realized. Overall, the TOP lessons ($M = 55.2$, $SD = 9.9$) were longer than the BAU-lessons ($M = 47.1$, $SD = 9.0$), although the duration of explicit instruction did not differ. BAU instruction focused on activating prior knowledge and, to a lesser

extent, on training reading strategies, without providing explicit text structure instruction.

TABLE 4

Realization and duration per lesson phase per condition

	Realized (% obs. lessons)		Duration (min)	
	TOP	BAU	TOP	BAU
Instruction	100	100	7.31 (3.28)	10.60 (8.03)
Modeling	95	30*	8.83 (3.05)	1.60 (1.99)*
Guided practice	75	70	6.27 (3.59)	10.80 (12.15)
Collaboration	100	40*	21.01 (8.93)	3.00 (4.52)*
Individual practice	70	80	9.00 (6.75)	19.80 (14.12)*

Note. TOP: intervention; BAU: business-as-usual. *Means differed significantly ($p < .05$).

Logbooks and lesson observations showed that teachers followed the intervention program quite strictly. However, there were issues related to conditional knowledge: in less than half of the observed intervention lessons (45%), teachers actually explained when a specific reading approach would be most useful. Also, instead of providing a moment of reflection on their reading approach after modeling, several teachers (30%) asked interactive questions during modeling. Logbook data show that teachers felt able to provide instruction ($M = 3.19$, $SD = 1.26$; on a five-point scale), but that they rated the lessons as somewhat difficult for their students ($M = 3.73$, $SD = 0.66$). During interviews, teachers mentioned that the instruction took longer than usual, and that some text topics were challenging.

Business-as-usual lessons put emphasis on individual student activities (42% of total lesson time), while the TOP program mainly focused on collaborative activities (38% of total lesson time) and much less on individual student activities (16% of total lesson time). Business-as-usual collaboration consisted of working in pairs on a few questions; only one of the observed teachers used another collaborative learning structure. By contrast, the TOP program prescribed many jigsaw collaborative tasks. Teachers were very positive about the collaborative

activities, but several implementation issues occurred: in half of the observed TOP lessons, teachers did not create expert and mixed groups, but turned the jigsaw tasks into simple pairwise activities, which might undermine the jigsaw principle of information exchange and positive interdependence between students. In addition, lesson observations and logbooks revealed that teachers often skipped the individual activities and feedback at the end of intervention lessons. This can be linked to interview data: teachers complained about lesson duration, and indicated that they regularly skipped individual activities to shorten lessons.

A comparison of the way in which the intervention was implemented in the TOP-1 group and the TOP-2 group, displayed some differences. Observation data show that relative to the other group, the TOP-2 teachers devoted 35% more time to explicit instruction ($\Delta M = 2.22$, $SE = 1.52$, $p = .04$) and 9% more time to individual practice ($\Delta M = 0.78$, $SE = 3.13$, $p = .03$).

Immediate and delayed effects

The following sections (see next pages) discuss the immediate and delayed intervention effects in both groups per outcome measure. Appendix D provides more details on the model fitting procedures, parameter estimates, and the estimated means.

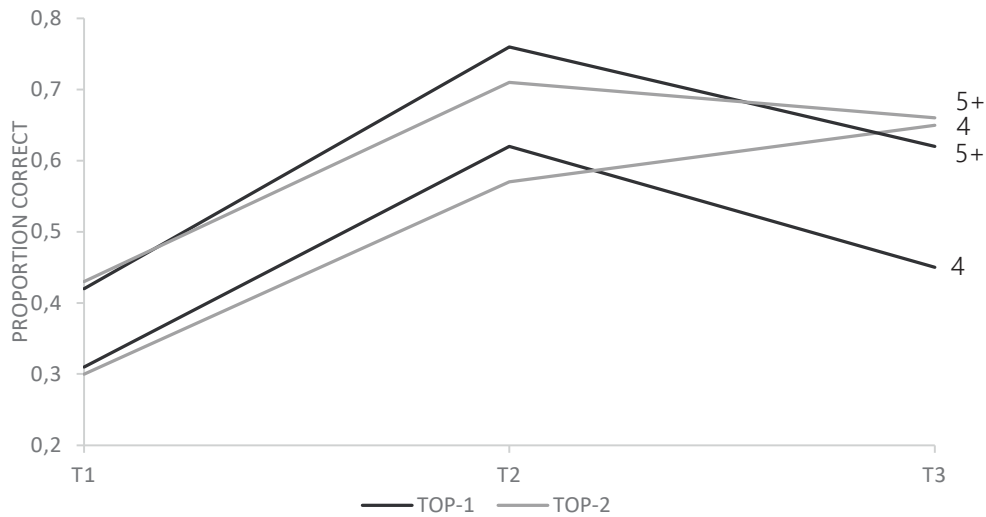
Text structure test

Figure 2 provides an overview of the estimated means per group and grade on the text structure test at three measurement occasions. There were main effects of Grade ($F(1, 10556) = 7.47, p = .006$) and Time ($F(2, 10556) = 3.40, p = .03$), and interaction effects of Time and Group ($F(2, 10556) = 21.34, p < .001$), Grade and Time ($F(2, 10556) = 3.11, p = .04$), and a three-way interaction of Grade, Group and Time ($F(2, 10556) = 4.62, p = .01$).

For the TOP-1 group, there was no demonstrable immediate and therefore no delayed effect of the intervention on the text structure test; at T2, the groups did not differ significantly (all $ps > .05$). For the TOP-2 group, there was an immediate effect of the intervention, with an effect size of $d = 0.50$, but only for fourth graders. Different from the other groups, they scored relatively higher at T3 with respect to T2, so that they outperformed fourth graders in the TOP-1 group ($\Delta M = .20, SE = .07, p = .006$), and scored comparable to older students at T3 ($\Delta M = .01, SE = .07, p = .87$).

FIGURE 2

Text structure test: Estimated means per grade and group



Standardized reading test

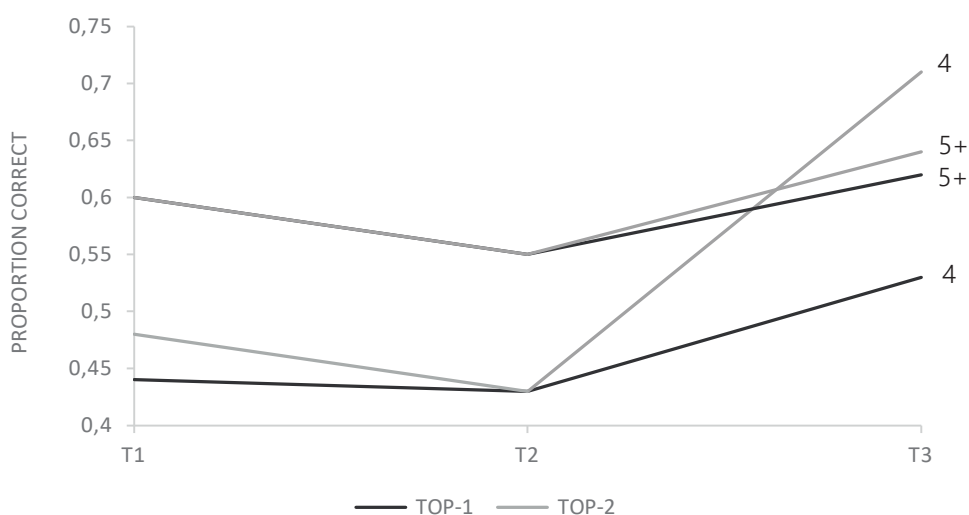
Figure 3 shows the estimated means for the standardized reading test at three measurement occasions (T1, T2, T3). There was a main effect of Grade ($F(1, 7418) = 4.30, p = .04$), as well as interaction effects of Time and Group ($F(2, 7418) = 4.53, p = .01$), Grade and Time ($F(2, 7418) = 9.81, p < .001$), and a three-way interaction between Grade, Time, and Group ($F(2, 7418) = 3.34, p = .04$).

For the TOP-1 group, no demonstrable immediate effect and as a consequence no delayed effect could be shown on the standardized test: neither of the groups improved between T1 and T2, so that both groups did not differ at T2 (all $ps > .05$).

For the TOP-2 group, we found an immediate effect of the intervention, but only for fourth graders, with an effect size of $d = 0.53$. Between T2 and T3 they significantly improved ($\Delta M = .28, SE = .17, p = .045$), so that at T3, they outperformed fourth graders in the TOP-1 group ($\Delta M = .18, SE = .07, p = .01$), and scored comparable to the fifth and sixth graders at T3 ($p = .98$). Overall, the variance between items is relatively high on the standardized reading test ($S^2 = 2.89, SE = .80$), suggesting that items might have been too heterogeneous to demonstrate effects.

FIGURE 3

Standardized reading test: Estimated means per grade and group



Summarization

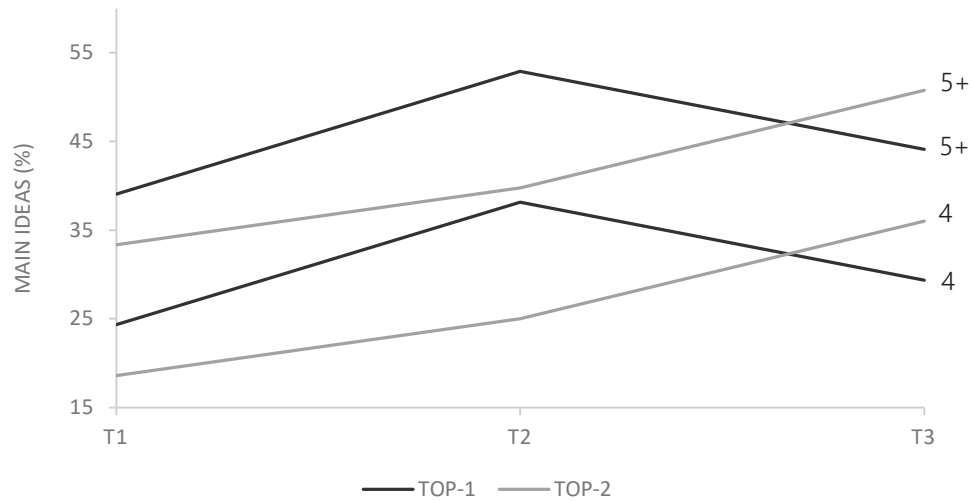
Figure 4 shows the estimated means for the summarization task at three measurement occasions. For summarization – and specifically the main ideas included in students’ summaries – we found a main effect of Grade ($F(1, 20) = 14.12, p = .001$); fourth graders included 14% fewer main ideas in their summaries than older students ($\beta = -.14, SE = 3.91$). In addition, a main effect of Time was found ($F(2, 553) = 13.90, p < .001$), and an interaction effect of Time and Group ($F(2, 553) = 9.38, p < .001$).

For the TOP-1 group, there was no immediate intervention effect, and as a consequence, no delayed effect. Even though TOP-1 students scored better on T2 with respect to T1 ($\Delta M = 13.82, SE = 3.07, p < .001$), the interaction effect of Time and Group between T1 and T2 was not significant ($\Delta M = -7.61, SE = 4.65, p = .10$).

For the TOP-2 group, there was an immediate effect of the intervention, with an effect size of $d = 0.48$. TOP-2 students scored better on T3 with respect to T2 ($\Delta M = 11.68, SE = 5.10, p = .005$), while the TOP-1 group scored lower ($\Delta M = -8.82, SE = 3.15, p = .005$). The interaction effect of Time and Group between T2 and T3 was significant ($\Delta M = 19.67, SE = 4.64, p < .001$).

FIGURE 4

Summarization: Estimated means per grade and group (% main ideas)



Metacognitive knowledge

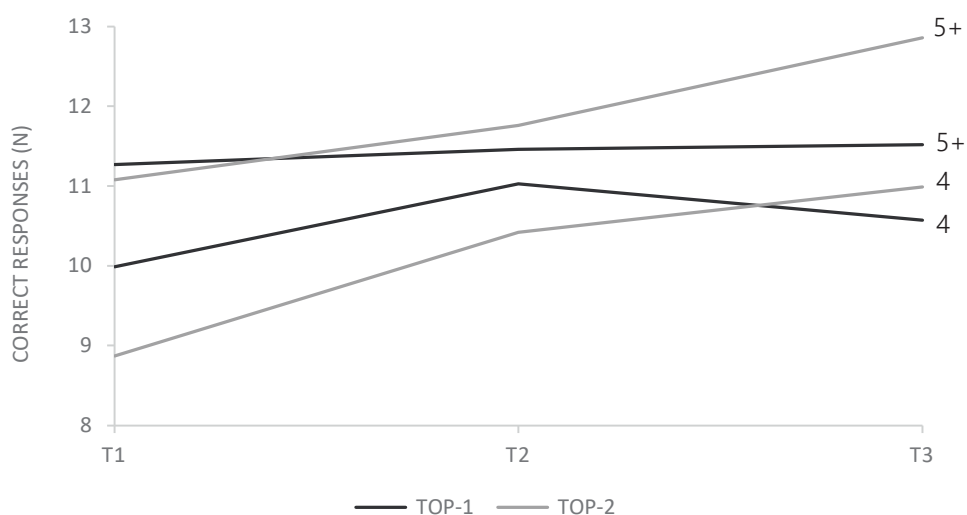
Figure 5 shows the estimated means for explicit metacognitive knowledge at three measurement occasions. There were main effects of Grade ($F(1, 22) = 21.61, p = .001$) and Time ($F(2, 13) = 12.47, p < .001$), as well as an interaction effect of Time and Group ($F(2, 572) = 5.39, p = .005$).

For the TOP-1 group, there was no immediate effect, and as a consequence, no delayed effect of the intervention. That is, the interaction effect of Time and Group between T1 and T2 was not significant ($\Delta M = 0.49, SE = 0.47, p = .30$). At T2, there were no significant differences between groups (all $ps > .05$).

For the TOP-2 group, there was an immediate effect of the intervention program, with an effect size of $d = 0.24$. The interaction effect of Time and Group between T2 and T3 was significant ($\Delta M = 1.03, SE = 0.48, p = .03$). The TOP-2 group scored better on T3 with respect to T2 ($\Delta M = 0.88, SE = 0.35, p = .01$), while the TOP-1 group remained stable ($p > .05$). As a result, the TOP-2 group outperformed the other group at T3 ($\Delta M = 0.92, SE = 0.41, p = .03$).

FIGURE 5

Metacognitive knowledge: Estimated means per grade and group (score 0-15)



Writing

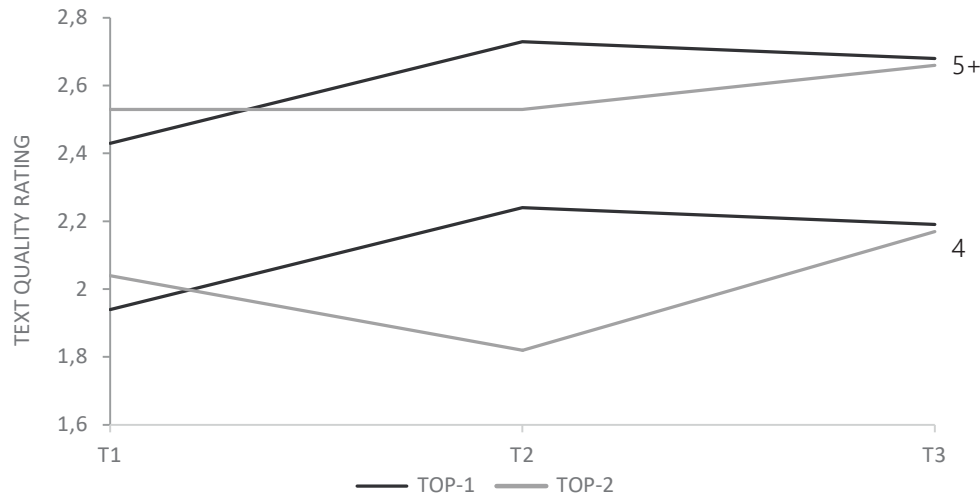
Figure 6 shows the estimated means for writing at three measurement occasions. A main effect of Grade could be demonstrated ($F(1, 9.4) = 31.48, p = .003$), showing that fourth graders scored lower than students in the other grades on writing ($\beta = -0.42, SE = 0.11, p < .001$). In addition, an interaction effect of Time and Group ($F(2, 561) = 4.31, p = .01$) was found.

For the TOP-1 group, there was an immediate effect of the intervention, with an effect size of $d = 0.33$. That is, students in this group scored higher on T2 than on T1 ($\Delta M = 0.30, SE = 0.12, p = .02$), while the TOP-2 group scored lower ($\Delta M = -0.22, SE = 0.14, p = .11$). As a result, TOP-1 students outperformed students in the regular curriculum at T2 ($\Delta M = 0.42, SE = 0.14, p = .004$). This interaction effect of Time and Group between T1 and T2 was significant ($\Delta M = 0.53, SE = 0.19, p = .005$).

For the TOP-2 group, we also found an immediate effect of $d = 0.39$. The TOP-2 group scored higher on T3 with respect to T2 ($\Delta M = 0.36, SE = 0.14, p = .009$), while students in the TOP-1 group remained quite stable in their writing between T2 and T3 ($p > .05$). This interaction effect of Time and Group between T2 and T3 was significant ($\Delta M = 0.40, SE = 0.18, p = .03$).

FIGURE 6

Writing: Estimated means per grade and group (rating 0-5)



5.4 Conclusion and Discussion

This quasi-experimental study evaluated the effectiveness of a text structure intervention (TOP) that was developed for Dutch students in the upper elementary grades. With a switching-panels replication design, we evaluated the intervention effects in two iterations (TOP-1 and TOP-2) on various measures of reading comprehension, on explicit metacognitive knowledge, and on writing.

For most outcome measures, one out of four possible immediate intervention effects could be demonstrated. That is, only fourth graders in the TOP-2 group outperformed other students on the text structure test ($d = 0.50$), the standardized reading test ($d = 0.53$), the summarization task ($d = 0.48$), and the explicit metacognitive knowledge test ($d = 0.24$), directly after finishing the intervention program. On the writing task, immediate effects were demonstrable across grades in both iterations of the intervention ($d = 0.33$ and $d = 0.39$).

The effects that were demonstrated for the fourth graders in the TOP-2 group resonate with meta-analytic research: text structure instruction can have a positive effect on text structure knowledge, summarization, and students' performance on reading comprehension tests (*Chapter 2*; Hebert et al., 2016; Pyle et al., 2017). Knowledge about text structure provides students with a tool to identify main ideas and organize these in a meaningful framework, which is crucial for text comprehension and for getting the gist of the text (Meyer et al., 1980; Miyatsu et al., 2018; Stevens & Vaughn, 2021; Strong, 2020; Taylor, 1985; Westby et al., 2010). For this reason, researchers have argued that text structure instruction should be a crucial component of reading and summarization instruction (Miyatsu et al., 2018; Stevens & Vaughn, 2021; Wijekumar et al., 2019), in particular as current textbooks for reading often display a poor coverage of text structures and summarization skills (Beerwinkle et al., 2018; *Chapter 3*; Wijekumar et al., 2019).

In both iterations of the intervention, there appeared to be an immediate transfer effect on students' writing skills. Students may have benefited from the explicit text structure instruction and the writing tasks in the TOP program. This is consistent with the assumption that knowledge of text structures provides readers with a useful frame to organize information in their mind, and writers with a means to organize information on paper (Dickson, 1999; Hebert et al., 2018; Strong, 2020). Our study complements previous research, as few merely reading-focused text structure interventions actually examined transfer effects on writing, probably

because spontaneous transfer effects from reading to writing are considered to be less common than the other way around (Couzijn, 1996). It would be worthwhile if more reading-focused text structure research would include writing measures to replicate this finding, and also to examine whether transfer effects are more pronounced on more aligned informational writing tasks.

In the current study, most findings are not generalizable, as the effects were often only demonstrated for one particular group of students: the fourth graders in the second intervention group. We should be cautious to claim that the text structure lessons had little effect: the group of students who showed a significant effect on the text structure test also showed effects on all other reading measures, suggesting that text structure knowledge can foster comprehension skills.

The remarkable outcome that only the fourth graders in one iteration of the intervention responded as expected to the intervention, might not have been discovered with a single quasi-experimental design. This illustrates the added value of the switching-panels replication design: it allows for studying the replicability of effects (Shadish et al. 2002), even though it might generate more questions than answers. Therefore, we will discuss our findings in the light of a critical inspection of test alignment and quality, intervention content, and fidelity of implementation.

Test alignment and quality

First, we need to reflect on the degree of alignment of the outcome measures with the intervention content. We intended to measure students' explicit knowledge about text structures and reading strategies. In hindsight, the explicit metacognitive knowledge test (Brand-Gruwel, 1996) lacked a good alignment with the intervention content (Clemens & Fuchs, 2021), and might be even more aligned with the current, business-as-usual Dutch reading curriculum (see *Chapter 3*) than with the intervention content. That is, the items focused on a general, non-contextualized use of many reading strategies, whereas our intervention program aimed to promote a contextualized insight in how a small set of reading strategies should be applied in specific text structures (Hoch & McNally, 2020; Stevens & Vaughn, 2021). This might partially explain why the effect size on this measure was relatively small compared to the other outcome measures.

It is remarkable that not all intervention groups performed better at the text structure test right after finishing the intervention program. The text structure test may have turned out less aligned with intervention content than intended. For example, several graphic organizer items were inspired by a study among secondary school students (Land, 2009), but these items did not perfectly resemble the graphic organizers of the intervention program. Studies with better alignment in this respect often get clearer results (e.g., Strong, 2020). In addition, several test items did not directly measure students' explicit text structure knowledge, but required multiple steps of reasoning. For example, instead of asking students which cue words can be found in cause-effect texts, or instead of having them fill in a cloze task, students were asked: 'Peter reads a text on the consequences of unemployment; what signaling words will he probably encounter in the text?' This requires students to infer in a first step that this text probably has a cause-effect structure, and in a second step, recall which signaling words are characteristic of cause-effect structures.

Hence, the text structure test turned out to be a mid-transfer test at best (Clemens & Fuchs, 2021). For future research, it therefore seems wise to develop a text structure test that is better aligned with intervention content, and to use a more indirect approach of measuring text structure knowledge, for example by using cloze tasks to measure structure-specific signaling word knowledge, and using (multiple-choice) questions related to main idea identification and inference making based on the text structure (see Wijekumar et al., 2017).

The lack of consistent outcomes on the standardized reading test across intervention groups also raises questions with regard to alignment. It probably reflects the fact that standardized reading tests measure more strategies and skills than those directly related to insight into text structures (Hebert et al., 2016; Leslie & Caldwell, 2009; van den Broek & Kremer, 2000). Our standardized reading tests also contained literal comprehension questions and questions pertaining to narrative texts, which might have made the test less sensitive to readers' situation model reading skills for informational texts; the skills that are assumed to be primarily affected by text structure instruction (Kintsch, 2004; Meyer & Ray, 2011; Wijekumar et al., 2019; Williams, 2018). Therefore, the standardized test should be considered as a far transfer test on which less consistent effects could be expected (Clemens & Fuchs, 2021).

Another issue that warrants a critical reflection is the quality of the tests in relation to the research design. In order to examine growth patterns and effect maintenance, a switching-panels replication design requires three perfectly parallel constructed tests for each measurement occasion (i.e., tests with equal means, observed and true score variances). If this condition is met, changes in scores between measurement occasions can unequivocally be interpreted as growth. As there were no Dutch parallel reading tests freely accessible, all reading tests had to be developed from scratch, and all tests needed to be perfectly parallel. Despite our attempts to create three parallel reading tests from a large item data base (Kuhlemeier et al., 2014), and to scale and pilot the researcher-designed text structure test in three non-participating schools, the tests did not turn out as perfectly parallel, and substantial between-item and error variance was detected.

Due to the lack of test parallelism, it was impossible to analyze growth patterns. For the analysis of intervention effects, we had to base our data analysis on interaction effects, instead of growth patterns. The power of a statistical analysis based on interaction effects is lower than one based on main effects of growth. In the light of a limited test quality, it might simply have been too ambitious to prove intervention effects in two iterations, over and above business-as-usual gains, while generalizing over classes, students, and items. Therefore, the construction of high-quality reading tests for the Dutch context – in particular with a focus on higher-order text comprehension – should be an important focus of future research, so that questions on growth, maintenance, and replicability can be answered. Moreover, the development of such tests might stimulate teachers as well to focus more on higher-order comprehension skills (van den Broek et al., 2021).

Intervention content

Second, the limited effectiveness of the text structure intervention also warrants a critical reflection with regard to the intervention duration and content. Text structure interventions with more favorable results in the US often had a higher intensity and/or intervention duration, such as fifty lessons during a whole schoolyear (Williams et al., 2016), four lessons a week during eight weeks (Strong, 2020), or one web-based lesson each week during seven months (Wijekumar et

al., 2017). In another study among Dutch primary school students, the effects of a reading intervention only showed up after two years of teaching (Droop et al., 2016). On the other hand, our intervention of ten hours was only marginally shorter than what Pyle and colleagues (2017) identified as the most favorable duration of text structure interventions (i.e., 11-20 hours). It is more likely that intervention duration in itself does not fully explain the limited results, but that the complexity of the intervention in relation to its duration should be considered (see also Hebert et al., 2016).

That is, the TOP intervention program was very different from the regular reading curriculum (see *Chapter 3*), in terms of both content (e.g., learning about four informational text structures, structure-specific strategy use), and instructional approach (e.g., collaborative learning activities). From this perspective, the duration of the intervention program might have been too ambitious: students had to learn about four text structures in ten one-hour lessons divided over six weeks.

Content-related aspects of the TOP intervention might have been less optimal as well. For example, the text structure intervention that was evaluated by Strong (2020) provided ample opportunities for students to gradually learn how to construct and complete text structure graphic organizers by themselves, as an active construction of graphic organizers is related to increased intervention effectiveness (*Chapter 2*). As we did not want to overload the lesson program of TOP, we might not have included a sufficient number of such tasks for students to practice on their own, and instead mainly offered them in the context of collaborative learning, which might have limited the effectiveness.

Implementation fidelity

We cannot pinpoint the exact reason for the fact that the fourth graders in the TOP-2 group were the only ones to demonstrate all of the hypothesized effects. It is not likely that the lack of effects in the other groups is due to social validity issues: all teachers were very positive about the intervention program and the opportunities it provided to integrate reading comprehension instruction with content-area subjects (see also Hwang et al., 2021; Maerten-Rivera et al., 2016). Zooming in on the teachers of the group who did demonstrate all the effects, it was found that they had on average less teaching experience, and spent slightly

more time on explicit instruction and individual practice than the other teachers. Generally speaking, young and less-experienced teachers rely more on teaching materials (e.g., Valencia et al., 2006), but this could be a good thing in case high-quality curricular materials are used. However, we should be very cautious in drawing strong conclusions, as the number of participating teachers is too limited.

Overall, teachers' implementation of the TOP lessons was not completely as intended, which was observed across teachers in all groups. This is not surprising, as many elements of the TOP program differed from the regular reading curriculum (*Chapter 3*), and one single training session may not have sufficed. Most issues were related to implementing a GRR: the collaborative learning task was often transformed into simple pairwise work, and individual activities were regularly skipped. This might have been affected by time management and scheduling issues (Gillies & Boyle, 2010; Hebert et al., 2018), as well as by limited teacher knowledge and skills related to collaborative learning (Hacker & Tenent, 2002; Okkinga et al., 2021; Spörer et al., 2009; Veldman et al., 2020).

All in all, the promising finding in one of the intervention groups suggests that text structure instruction could be a promising approach to foster reading-to-learn skills in the upper elementary grades, but various challenges still have to be overcome in the Dutch context. Although it is tempting to suggest that the lack of a GRR has tempered the effectiveness of TOP, the implementation data are too limited. Therefore, the only solid conclusion at this point is that we need to gain more insight into implementation fidelity by examining the challenges related to implementing a GRR and/or text structure instruction. This can inform the design of interventions and/or curricular materials that can support teachers' application of effective instructional approaches during text structure instruction (Beerwinkle et al., 2018; Hebert et al., 2018).

CHAPTER 6



A script for success? Teachers' fidelity of implementation of text structure lessons for primary education

Abstract

The success or failure of reading interventions depend partly on teachers' implementation in the classroom. Despite the limited amount of research on the nature of these implementation issues, various studies suggest a relation to the implementation of a gradual release of responsibility (GRR). The current multiple-case study foregrounds how three teachers enacted a scripted reading intervention program focused on informational text structures (TOP), in which instruction and activities followed a GRR. Based on logbooks, lesson observations, and interviews, we describe which implementation challenges arose during each phase of the GRR, by analyzing both structural dimensions (e.g., timing and duration of activities) and process-oriented dimensions (e.g., instructional techniques, feedback practices). Implementation fidelity scores varied across teachers and lesson phases, with the lowest scores for collaborative learning (51%) and the highest for explicit instruction (91%). No overall positive correlation was found between teachers' satisfaction with lesson phases and the fidelity of implementation scores these phases obtained. Many implementation issues were related to (1) promoting metacognitive knowledge, (2) promoting effective collaboration, and (3) a tension between limited teacher knowledge and teacher autonomy. Students' on-task behavior during collaborative tasks seemed related to teachers' skillfulness in implementing such activities. These findings suggest that teachers need more knowledge and support related to promoting metacognitive knowledge and guiding collaborative learning in whole-classroom settings. This study may inspire the design of teaching materials that are not too tight but provide adequate teacher support.

6.1 Introduction

When children read informational texts, they often engage in superficial text processing (Diakidoy et al., 2003; Helder et al., 2016; Kim & van Dusen, 1998; Rooijackers et al., 2020), and struggle to filter main ideas and generate accurate inferences (Kraal et al., 2018; Wijekumar et al., 2020). One of the reasons why informational texts are notoriously difficult, is the fact that they can have different underlying text structures (Coté et al., 1998; Schleppegrell, 2004), such as compare-contrast, cause-effect, problem-solution, chronology, or combinations thereof (Meyer, 1975), which puts demands on the specific way in which readers need to make inferences and connections (Cromley & Azevedo, 2007; Hiebert & Mesmer, 2014; Williams et al., 2004). For instance, the creation of a coherent mental representation of cause-effect texts requires another approach than that of compare-contrast texts: readers should make different predictions and inferences (i.e. infer causal relationships vs. comparative relationships), and take different steps to reach a coherent summary or visualization (e.g., cause-effect schemas vs. Venn diagrams) (see Hoch & McNally, 2020; Stevens & Vaughn, 2021).

A large body of research has shown the importance of explicit text structure instruction as a way to improve students' reading comprehension (Hebert et al., 2016; Meyer & Ray, 2011; Pyle et al., 2017), also for students in the upper elementary grades (*Chapter 2*). Explicit text structure instruction helps students to recognize the underlying text structure of narrative and informational texts, supports students' inferencing, and fosters their summarization skills (Miyatsu et al., 2018; Stevens & Vaughn, 2021; Westby et al., 2010; Wijekumar et al., 2021). Yet, teacher quality seems to be a decisive factor in the ultimate success of reading comprehension instruction (Beerwinkle et al., 2018; Connor et al., 2014).

For many years, research has focused on identifying what interventions work, but much less attention has been devoted to the specific strengths and weaknesses of interventions and the conditions for success: the why and how (Century et al., 2010). As a consequence, relatively little is known about what teachers precisely do in their classroom while providing text structure instruction, and how they enact their teaching materials (but see Beerwinkle et al., 2018). Zooming in on the fidelity of implementation – that is, examining the extent to which critical components of an intervention are present when that program is

enacted (Century et al., 2010) – therefore could help to interpret the findings from intervention studies (Durlak & DuPre, 2008).

A study on fidelity of implementation is particularly relevant in the context of text structure instruction, as the intervention success varies from study to study (*Chapter 2*; Pyle et al., 2017) and can even vary across experimental groups within the same study (see *Chapter 5*). Also, generally speaking, reading interventions are often more effective when implemented by researchers instead of teachers (Dignath & Buettner, 2008; Okkinga et al., 2018a; Pyle et al., 2017). This raises questions related to implementation fidelity and, maybe even more importantly, what specific knowledge and support teachers need for successful implementation (Beerwinkle et al., 2018; Bohaty et al., 2015; Strong, 2020; Williams, 2018). Therefore, this chapter details how teachers actually enact an intervention program focused on text structure instruction.

Research on implementation challenges related to reading instruction is not completely new. Various factors appear to be related to variations in the implementation of interventions. For example, factors such as available instruction time and class management issues often explain differences in implementation fidelity, mainly with respect to the completion and duration of program components (Gillies & Boyle, 2010), which is referred to as *structural dimensions* of implementation fidelity (Century et al., 2010; Harn et al., 2013). Factors related to the nature and quality of teacher-student interactions during interventions are referred to as *process-oriented dimensions* (Century et al., 2010; Harn et al., 2013).

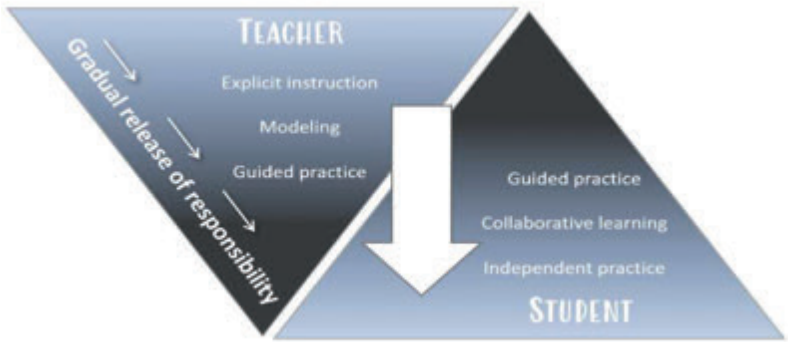
The structural and process-oriented dimensions of implementation fidelity seem to affect intervention success and student literacy outcomes (Connor et al., 2014; Odom et al., 2009; Taylor et al., 2003), and may be related to teachers' pedagogical content knowledge (Ball & Cohen, 1996; Century et al., 2010; Gudmundsdottir & Shulman, 1987; Penuel et al., 2014; Valencia et al., 2006). This is especially relevant in the current study, given reports on limited teacher knowledge about text structures (Beerwinkle et al., 2018; Reutzel et al., 2016).

It has been suggested that – for an optimal instructional approach during reading comprehension lessons – the teacher should first provide explicit instruction and modelling, and then gradually step back and transfer the responsibility for learning to students with decreasing levels of scaffolding; first through guided practice and teacher feedback (Nolte & Singer, 1985; Pressley et

al., 1992), and later on through collaborative small-group learning (Brown 1984; Fuchs et al., 1997; Klingner & Vaughn 1999; Palinscar & Brown, 1984), until students are capable of practicing on their own. These insights have empirical back-up and are combined in the Gradual Release of Responsibility model, as illustrated in Figure 1 (GRR model; Fisher & Frey, 2021; Kelly, 2019; Pearson & Gallagher, 1983).

FIGURE 1

Overview of the GRR model: the triangles show the proportion of responsibility divided between teacher and students per lesson phase.



Yet, the implementation of a GRR model might be challenging. For example, research has shown that teachers’ limited knowledge about modeling and collaborative learning can hinder a successful enactment of reading interventions (Okkinga et al., 2018b; 2021; Spörer et al., 2009; Veldman et al., 2020). Moreover, teachers often struggle to progress in a very gradual way from teacher-centered to student-centered activities (Hacker & Tenent, 2002), and do not always have the skills and knowledge to effectively guide meaningful collaborative learning in classroom settings (Hacker & Tenent, 2002; Okkinga et al., 2018b; Spörer et al., 2009). This might especially be the case with young students, as these possess limited cognitive and metacognitive skills for completing collaborative tasks (De Backer et al., 2021).

In the current multiple-case study, we focus on the implementation of a text structure intervention (TOP; see next section) in which the instruction and activities follow the GRR model. The intervention was initially developed in close collaboration with four co-designing teachers (*Chapter 4*) and a revised version

was then tested in a large-scale effect study (*Chapter 5*). The intervention appeared successful on various outcome measures, but only in one of the intervention groups. This raised questions on implementation fidelity: why would the intervention work in one of the groups, but not in the others? Limited implementation data suggested that several teachers skipped individual practice, or transformed small-group collaborative learning activities into pairwise work, but it was impossible to pinpoint the exact reasons for the variability in outcomes.

As our study is not the only one suggesting that there are often implementation issues related to the implementation of a GRR (e.g., Hacker & Tenent, 2002; Okkinga et al., 2018b; 2021; Spörer et al., 2009; Veldman et al., 2020), the current multiple-case study aims to examine to what extent teachers' adaptations in a text structure intervention can hinder the realization of a successful GRR. The following questions guided the current study:

- RQ1: Which structural and process-oriented implementation challenges arise during the different phases of the GRR model when primary school teachers enact a text structure intervention?
- RQ2: To what extent is teachers' implementation fidelity related to their satisfaction with each phase of the GRR, and to their students' overall satisfaction and behavior?

This will help clarify the conditions for a successful implementation of text structure instruction and/or the GRR model, and can identify specific aspects that are difficult for teachers to implement (Harn et al., 2013). Moreover, a close inspection of the way in which teachers adapt the intervention to their own classroom environment can generate useful insights for the future design of interventions that are sensitive to teachers' professional development needs (Brown, 1992; Lindo & Elleman, 2010; Snodgrass et al., 2016). Uncovering implementation challenges might form a first step towards the design of so-called educative curriculum materials: curricular materials that are designed to simultaneously address teacher and student learning by offering additional content for teachers, so that they can also grow in their pedagogical content knowledge (Ball & Cohen, 1996; Davis & Krajcik, 2005; Davis et al., 2017; Edelson et al., 2021; Haas et al., 2021; Schneider et al., 2005).

TOP intervention: content and instructional approach

The TOP intervention provides explicit instruction on informational text structures for students in the upper elementary grades (Table 1). It consists of eight one-hour lessons on four different text structures: compare-contrast (Lesson 1 & 2), problem-solution (Lesson 3 & 4), chronology (Lesson 5 & 6) and cause-effect (Lesson 7 & 8), as these structures may be just within the zone of proximal development for students at this age (Meyer et al., 2018). Students receive instruction on how to recognize these text structures, and how to apply several reading strategies – predicting, questioning, visualizing, and summarizing – in a way that fits the specific text structure at hand.

Text structure often forms a difficult topic for primary school teachers (Beerwinkle et al., 2018; Reutzel et al., 2016), and Dutch teachers receive only limited instruction about it on teacher training colleges (Kooiker-den Boer et al., 2019). Therefore, the intervention program was highly scripted: all instruction and activities were well described in the teaching manual, and PowerPoint slides guided teachers through each lesson. The manual also contained a theoretical overview of all structures. Four introductory video clips were created for teachers, which explained the four text structures and the guiding design principles.

Lesson 1 briefly introduced students to the concept of text structure and then went on to explain the main features of the compare-contrast structure. Lesson 2 focused on using text structure to summarize compare-contrast text in a graphic-organizer (Venn diagram), summarize the main idea, and answer some questions that relate to situation-model understanding of the text (e.g. *What is the main similarity between x and y?*).

The other lessons followed the same pattern: for each text structure, the first lesson focused primarily on text structure recognition in shorter text fragments, while the second lesson typically featured longer texts and offered instruction and practice with structure-specific summarization skills (e.g., identifying main ideas, using graphic organizers), as this seems to be a powerful way to raise students reading comprehension (*Chapter 2*; Miyatsu et al., 2018; Stevens & Vaughn, 2021). All texts were related to content-area subjects in order to promote transfer of reading-to-learn skills (Chan et al., 1992; Read et al., 2008).

The instruction and activities of all lessons followed the GRR model (Fisher & Frey, 2021; Pearson & Gallagher, 1983; Webb et al., 2019). Lessons started with

teacher-centered activities. First, during explicit instruction, the teacher introduced or reviewed one text structure. This was followed by a modeling phase in which teachers demonstrated how the text structure could be recognized, and how reading strategies (e.g., predicting, summarizing) could be applied in that specific text structure.

The teacher manual provided elaborate scripts for this modeling phase, in order to support teachers in their demonstration of a structure-specific use of reading strategies, and to make sure that teachers would not be tempted to engage in interaction *during* modeling, as this might put students in a role of participants instead of observers, which would challenge the foundational principles of observational learning. That is, modeling presupposes that simply observing the actions of a model already provides an opportunity to learn (Bandura, 1986). This key feature of observational learning separates *doing* from *learning how to do*, which is likely to focus students' attention on the process-related aspects of cognitive tasks (Zimmerman & Risemberg, 1997) and might reduce the cognitive load, as has been argued in the area of writing instruction (Evers-Vermeul & van den Bergh, 2009; Rijlaarsdam et al., 2005).





Right after teacher modeling, students and teachers were prompted to engage in a moment of reflection or abstraction (Schutz & Rainey, 2019), where they discussed their teachers' reading approach (e.g., *What strategies did I apply during reading, and why?*). Such an active reflection after modeling explicitly focuses students' attention at crucial aspects of the reading approach, which seems to be related to improved student outcomes, at least in the area of writing (Braaksma et al., 2002; Rijlaarsdam et al., 2005). In one lesson, students watched a video clip of a model who made mistakes, and were asked to explicitly compare this to the approach of their teacher who modeled the same fragment. Such a comparison task can benefit student learning (Gentner & Namy, 1999).

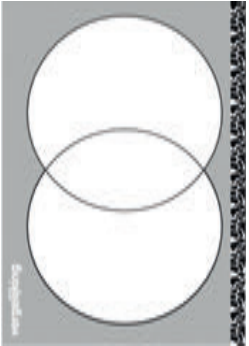
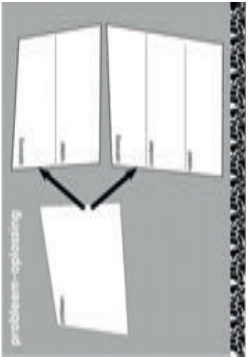
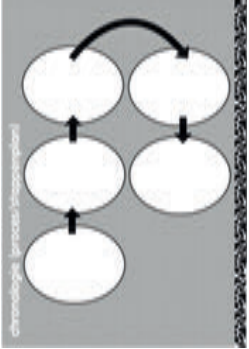
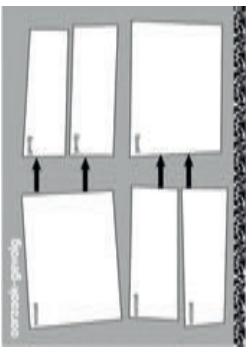
Then, the role of the teacher was gradually faded-out: after guided practice, lessons proceeded with student-centered activities. This started with small-group collaboration tasks, which were often jigsaw activities (Aronson, 1978). In a jigsaw reading task, students first work in expert groups (AAA, BBB, CCC) to make sense of their part of the text. Then, in a second step, they form mixed groups (ABC) to exchange information with students who read another (part of the) text. This information exchange is crucial to finally complete a joint product (Colosi, & Zales,

1998; Doymus,, 2007), which could be a completed graphic organizer, an information poster, or a preparation for a quiz for which information of all text parts matters. This type of collaborative learning can promote high-quality student interactions, and creates simultaneously individual accountability and positive interdependence between group members (Johnson & Johnson, 1984).

In the final step of the intervention lessons, students worked on an individual activity, as independent practice seems crucial for maintenance of text structure knowledge (*Chapter 2*). This often concerned some structure-based questions about the text (i.e., *What solutions does the text give to cope with El Niño?*), a metacognitive reflection task (e.g., *Can you provide tips for creating a good Venn diagram?*), or a writing task (i.e., writing a paragraph in a specific structure). During lesson closure, teachers and students reflected on the learning goals and imagined new situations where they could apply their knowledge about text structures (e.g. *Can you think of other situations where you have to read chronological texts?*) in order to explicitly promote transfer (Patton et al., 2022).

TABLE 1
Intervention program

Compare-contrast structure (CC) Lessons 1 & 2		Problem-solution structure (PS) Lessons 3 & 4		Chronological structure (CS) Lessons 5 & 6		Cause-effect structure (CE) Lesson 7 & 8	
Lesson 1: Instruction Text structure Recognizing CC	Lesson 3: Instruction Recognizing PS Introduction to PS-chart	Lesson 5: Instruction Recognizing CS (procedure) Using a flow chart	Lesson 7: Instruction Causes versus effects Recognizing CE and CE-chart (simple)				
<u>Texts</u> Coniferous or deciduous trees; Seeds and spores; Breathing; Jaguars or cheetahs?; North or South Pole?	<u>Texts</u> Living without a fridge; Invention of the train; Roman aqueducts; Roman underfloor heating; Roman currency	<u>Texts</u> Chocolate mousse; Cocoa or coffee? (CC); Harvesting cocoa; From Africa to the Netherlands; From bean to chocolate	<u>Texts</u> Extreme weather; Climate change in the Himalayas; Altitude sickness; The Yeti; Himalaya tourism				
<u>Activities</u> Structure recognition; Matrix task	<u>Activities</u> Discussion task; Questions	<u>Activities</u> Summary and quiz; Write a recipe	<u>Activities</u> Structure recognition; Summary				

<p><u>Lesson 2: Instruction</u></p> <p>Summarizing with a Venn diagram</p> <p>Matrix question</p> <p><u>Texts</u></p> <p>Mustelids; Rodents; Pack animals</p> <p><u>Activities</u></p> <p>Venn diagram; Writing task</p>	<p><u>Lesson 4: Instruction</u></p> <p>Summarizing with a PS chart</p> <p><u>Texts</u></p> <p>Ocean Clean-up: various solutions</p> <p><u>Activities</u></p> <p>Problem-solution chart</p> <p>Invent a solution (write PS text)</p>	<p><u>Lesson 6: Instruction</u></p> <p>Recognizing chronology (timeline)</p> <p>Summarizing with a timeline</p> <p><u>Text</u></p> <p>History of Utrecht</p> <p><u>Activities</u></p> <p>Create timeline</p> <p>Write letter</p>	<p><u>Lesson 8: Instruction</u></p> <p>Direct and indirect effects</p> <p>CE-charts (complex)</p> <p><u>Texts</u></p> <p>Impact of El Niño (Peru; Australia); El Niño or La Niña? (CC); Being prepared for El Niño (PS)</p> <p><u>Activities</u></p> <p>Information poster</p> <p>Quiz</p>	   
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

6.2 Method

Participants and research design

55 regular primary schools were approached for participation, and eventually three schools agreed to participate, each with one fourth grade teacher and their students ($N = 63$; 48% male).⁴ All schools were of a Protestant denomination, located in the middle of the Netherlands; one in a city, two in smaller towns. Due to ethics regulations, no information on special educational needs were collected. The three participating teachers (2 male, 1 female) had on average 10.3 years of teaching experience (min = 4, max = 21). For their business-as-usual instruction, they used teaching materials specifically designed to teach reading comprehension as a separate subject, and that mainly focus on reading strategies, but generally lack explicit instruction about informational text structures (see also *Chapter 3*). Teacher 1 had the least teaching experience. He taught 16 students and used *Nieuwsbegrip* [Understanding news]. Teacher 2 had the longest teaching experience and taught 24 students for which she used *Lezen in Beeld* [Reading in the picture], sometimes supplemented with *Nieuwsbegrip* or other materials. Teacher 3 taught a class of 23 students. He also used *Nieuwsbegrip* and was the reading coordinator at his school. All teachers were willing to replace their textbook for five weeks with the eight one-hour lessons of the TOP intervention. The lessons were completed at a rate of one or two lessons per week, during which implementation fidelity data were collected, through observations, logbooks, a student questionnaire, and teacher interviews.

Measures

Lesson observations

For each of the participating teachers, five lesson observations were scheduled during the implementation of TOP, as research shows that a minimum of five lesson observations is required for a valid impression of teachers' enactment (Heidelberg et al., 1993). The first and final lesson were not observed because

⁴ This study was approved by the Faculty Ethics Assessment Committee of Humanities (FETC-GW); study approval registered under number 21-186-03.

these lessons had a slightly different set-up, but each of the remaining lessons was observed in at least two different classrooms. One lesson observation could not be scheduled, so we ended up with 14 lesson observations.

The focus of lesson observations was both the structural dimensions of implementation fidelity (i.e., time allocation and completion) as well as the process-related dimensions (i.e., quality of the implementation of certain aspects of the program) (see Harn et al., 2013). An observation sheet (Appendix E) was used to evaluate for each lesson phase its duration, student involvement (2: high, 1: moderate, 0: poor), and an implementation check of key components (1: implemented as intended; 0: not implemented as intended). For example, for collaborative learning, it was checked whether the teacher first created expert and then mixed groups (i.e., according to the jigsaw principles), and whether the teacher provided feedback during the task. Observers were encouraged to provide an overall impression of the lesson and to explain their answers if something was not implemented as intended.

In addition, student-level observations were conducted by following the behavior of two randomly selected pairs of students ($n = 28$) during the collaborative learning activity. Appendix E shows the behavior that was checked (1: yes, 0: no) during this observation (e.g., Students ask for teacher support; Students show on-task behavior most of the time). In most lessons, this collaborative learning activity concerned a jigsaw task (Aronson, 1978) where students first worked together in an expert group and then had to exchange information in a mixed group in order to complete a joint product. Student behavior was checked once during the first part (expert groups) and second part (mixed groups) of the collaborative learning task.

All authors were involved in the design of the observation scheme. Some parts of the observation scheme had been used in previous studies to observe business-as-usual reading comprehension lessons (*Chapter 3*) or the previous version of the TOP intervention (*Chapter 5*). In order to check the validity of the observation scheme, three lessons were observed by two researchers. This resulted in an interobserver reliability of 90% for observations of student involvement, 94% for implementation aspects per lesson phase, and a correlation of $r = .99$ for the duration of lesson phases.

Logbooks

After each lesson, teachers filled out an online logbook (Appendix E). The first part of this logbook focused on the structural dimensions of implementation fidelity: the completion and duration of different lesson phases (implemented as intended, shorter than intended, longer than intended, omitted). The second part focused on teachers' implementation and satisfaction with teacher-centered activities, by asking them to rate different lesson phases on a 1-5 scale and to explain their answers in an open-ended question. The third part of the logbook was similar, but focused on teachers' implementation and satisfaction with student-centered activities (i.e., collaborative learning and individual practice). In the final part of the logbook, teachers were asked to rate the lesson as a whole, the texts, and provide open-ended feedback on the lesson. If teachers did not complete the logbook within the first two days after giving the lesson, a reminder was sent by email to complete the logbook as soon as possible. All teachers completed the logbooks for all lessons.

Interviews

In the week after the final lesson and third measurement occasion, semi-structured retrospective interviews were scheduled with each teacher. All interviews were audiotaped and transcribed by the first author. Questions were asked about teachers' experiences and satisfaction with the texts, the lesson content, and the teacher-centered and student-centered activities (Appendix E). Teachers were also interviewed about their motives for deviating from the teaching manual, and to explicitly compare the lesson series to their business-as-usual practice.

Questionnaire

A student questionnaire was printed on the final pages of the student workbooks. On a five-point scale – accompanied with happy, neutral, and sad emojis – students rated different statements related to their appreciation of the lesson series (e.g., *I liked working together in groups*, *I think that learning about text structure is useful*). In addition, students had to answer two open-ended questions in which they had to indicate their favorite and least preferred lesson. The teaching manual of lesson 8 asked teachers to have their students complete this

questionnaire at the end of the lesson series, yet only the students from one teacher completed it ($n = 25$).

6.3 Results

Structural dimensions of implementation fidelity

The overall level of agreement between the teachers' logbooks and the actual observed duration of lesson phases was $r = .61$. Not all teachers gave a perfect impression of their program adherence in their logbook: although for Teacher 2 and 3 the logbook reports mainly converged with lesson observations ($r = .68, p < .001$ and $r = .63, p = .001$), there was no significant correlation between logbook reports and lesson observations for Teacher 1 ($r = .23, p = .22$). Therefore, most findings in this section are based on lesson observations.

In almost all observed lessons, teachers realized all phases of the GRR model. Teacher logbooks and lesson observations both showed that the timing of half of the lesson phases was exactly as intended, that approximately one third of the lesson phases took shorter to complete, and that the remaining lesson phases took slightly longer to complete. In line with the directions in the teaching manual, most time was devoted to collaborative learning, explicit instruction, and individual practice. The phase of collaborative learning often took longer to complete, partially due to classroom management issues related to the grouping of students. As a result, teachers could often pay limited time to lesson closures: instead of the suggested closure with attention to transfer, lesson closure was a quick wrap-up of the lesson.

Table 2 provides an overview of the duration of lesson phases for each teacher, and shows the average suggested time in the teaching manual. The average observed duration was 58 minutes ($SD = 4.78$), which aligns with the average prescribed duration in the teaching manual of 58 minutes ($SD = 2.58$).

TABLE 2

Suggested and observed lesson duration in minutes (and standard deviations)

	Suggested duration	Observed lesson duration ($n = 14$)			
		Overall	Teacher 1	Teacher 2	Teacher 3
Explicit instruction	8.33 (2.58)	11.36 (4.78)	8.60 (4.7)	13.20 (5.93)	12.50 (1.29)
Modeling	8.33 (2.58)	6.71 (3.24)	4.60 (1.34)	9.20 (3.90)	6.25 (2.22)
Guided practice	5.83 (3.76)	7.08 (4.10)	3.25 (0.50)	9.60 (3.85)	8.00 (4.00)
Collaborative learning	18.33 (4.08)	22.36 (6.81)	19.40 (5.68)	23.40 (8.68)	24.75 (5.80)
Individual practice	11.67 (2.58)	11.21 (5.67)	8.40 (3.13)	14.40 (7.57)	10.75 (4.43)
Lesson closure	5.83 (2.58)	1.13 (1.35)	1.20 (1.67)	1.20 (0.45)	1.00 (1.35)
Total duration	58.33 (2.58)	58.93 (4.78)	45.00 (1.73)	70.80 (12.93)	61.50 (5.45)

There were considerable differences across teachers. Teacher 1 completed all lessons in a relatively short time with an average of 45 minutes ($SD = 1.73$) and had the highest rate of shortened or omitted lesson phases (47%), which mainly affected individual practice and the reflection on modeling. Teacher 2 realized an average lesson duration of 71 minutes ($SD = 12.93$), which is 13 minutes longer than suggested in the manual. She mainly deviated from the suggested time during explicit instruction and guided practice, as she added many anecdotes and interactive elements to it (e.g., mini whiteboards). In order to make up for this additional time, she systematically omitted lesson closures, as well as the reflection after modeling (60% of lessons).

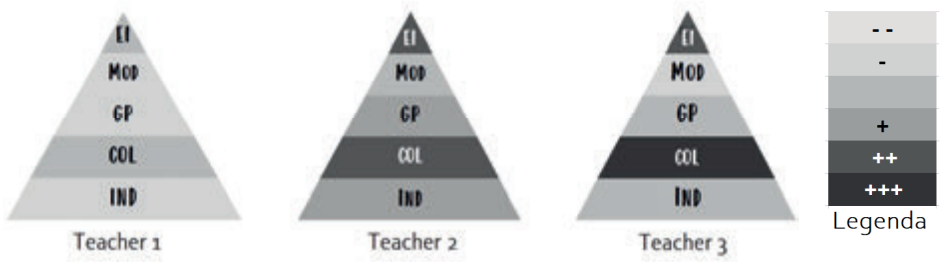
Teacher 3 respected the prescribed lesson duration with 60 minutes ($SD = 5.45$), yet he implemented less than half of the lesson phases in line with the suggested duration, both according to logbook data (39%) and observations (42%). He devoted more time to collaborative activities, and stated in his logbook: "I intentionally put a lot of emphasis on modeling, and on discussing my choices

with my students.” Yet, Teacher 3 realized the reflection after modeling in half of the observed lessons.

Throughout the lesson series, the teachers quite consistently deviated from the suggested timing of lesson phases. As this led to systematic emphasis or downplay of certain lesson phases, a GRR-pattern visualization was established for each teacher by calculating their mean deviation from the suggested lesson time per lesson phase (Figure 2). The darker the color, the more time was devoted to a lesson phase. The triangles represent the degree of student responsibility within the GRR that gradually expands when progressing from teacher-centered to student-centered activities. Figure 2 shows a relatively balanced pattern for Teacher 1, although the color intensity is relatively low due to the short duration of each lesson phase. For Teacher 2 and Teacher 3, the pattern shows emphasis on explicit instruction and collaboration. Teacher 2 also devoted more time to guided practice and individual practice.

FIGURE 2

GRR patterns: the darker the color, the more time was devoted to a lesson phase



Note. EI: Explicit Instruction; MOD: Modeling; GP: Guided Practice; COL: Collaborative learning; IND: Individual practice.
Legend: -- : 4 or more minutes shorter; - : 2 to 4 minutes shorter; + : 2 to 4 minutes longer; ++ : 4 to 6 minutes longer; +++ : more than 6 minutes longer.

Process-oriented dimensions of fidelity: implementation issues

This section discusses the implementation of the different phases of the GRR model and the implementation issues in each phase, mainly based on lesson observations.

Explicit instruction

On average, the teachers devoted 19% of their lesson time to explicit instruction. The teachers provided explicit instruction in accordance with the teaching manual, resulting in a high average implementation fidelity score ($M = 91\%$, $SD = 15$), and even a 100% score for Teacher 3 (Table 3). All teachers added elements to the explicit instruction (e.g., anecdotes, examples) to make it more engaging. Teacher 2 and 3 also prompted their students to use their mini whiteboards, for example to draw a timeline, or to note all the comparative signaling words they could remember. This worked well, but sometimes resulted in time management issues later on in the lesson.

TABLE 3

Percentage of lessons with explicit instruction as intended (and *SDs*)

Item	Overall	T1 (<i>n</i> = 5)	T2 (<i>n</i> = 5)	T3 (<i>n</i> = 4)
Teacher discusses lesson goals	100 (0)	100 (0)	100 (0)	100 (0)
Teacher focuses on text structure	86 (36)	80 (45)	80 (45)	100 (0)
Teacher provides clear instruction	79 (43)	60 (55)	80 (45)	100 (0)
Teacher reviews the previous lesson(s)	93 (27)	80 (45)	100 (0)	100 (0)
Total	91 (15)	84 (17)	92 (18)	100 (0)

Note. T1: Teacher 1; T2: Teacher 2; T3: Teacher 3

Looking at theoretical aspects that were problematic during this phase, all teachers mentioned that the content of Lesson 8 was difficult. In this second lesson on cause-effect structure, teachers were prompted to explain the distinction between direct and indirect effects. Also, the introduction of the problem-solution structure (Lesson 3 & 4) turned out to be a challenging one: teachers primarily focused on positive and negative adjectives, but not on connectives (e.g., *in order to*, *so that*) indicating a solution in a more subtle way.

Teacher 1 clearly struggled with the problem-solution structure: he could not explain the difference between problems and disadvantages. Different from what was prompted in the teaching manual, he gave the wrong instruction that disadvantages are part of the problem and that advantages are part of the solution, while advantages and disadvantages are in fact evaluative aspects relative to the solution. By contrast, Teacher 2 had a good understanding of problem-solution structure, and even added information to the lessons by making her students aware of their intonation as a cue for recognizing disadvantages (1).

1) A disadvantage is something less fun, right? You can tell from the tone of your voice. Listen! [*Reads aloud with exaggerated intonation:*] "Admittedly, the disadvantage is that (...)"

Did you hear that, guys? That intonation is different from when you read an advantage: then you read with your happy voice, not with a sad one. Let's read it together now and see if you can hear it too in your own voice! (Observation, T2)

Modeling

The average observed fidelity of implementation was relatively low ($M = 62\%$) for modeling (Table 4), in particular for Teacher 1 (51%). One issue during this phase was that teachers sometimes simply read the modeling script aloud from the teaching manual. Where the modeling scripts helped teachers to keep the focus of their think aloud on text-structural features, the scripts might have been too detailed. For example, Teacher 2 acknowledged that the scripts helped her prepare her lessons and focus her modeling on text structure, but she also believed that the scripts could make instruction too artificial (2).

2) Sometimes, modeling and reflection feels a bit farfetched and a little bit fake, because it is all scripted. I model with arithmetic as well, but then I do not have an example script and it feels more my own. I don't know how it came across to the children. (Interview, T2)

Teacher 2 reported that, throughout the lessons, she more and more felt that she could vary on the modeling scripts. However, lesson observations also revealed a

downside to improvisational modeling: once Teacher 1 and 2 started improvising, they also tended to engage more frequently in interactive discussions with their students while thinking aloud, even though the manual prompted teachers to ask students interactive questions *after* and not *during* modeling.

One prominent implementation challenge across teachers concerned the metacognitive reflection: teachers often omitted the reflection after modeling, and if they realized one, this reflection often lacked emphasis on metacognitive aspects of the reading process. In addition, teacher 3 commented on the reflection task in Lesson 2 where students explicitly had to compare the perfect example set by their teacher to a videoclip of a coping model who made some mistakes while reading a compare-contrast text, some of which were repaired. Teacher 3 felt that showing a coping model was not right and that a mastery model was needed to set a perfectly clear example, as shown in (3).

3) The wrong format was used for modeling: You need to show the right way once you teach something new, not to demonstrate what can go wrong. (Logbook, T3)

TABLE 4

Percentage of lessons with teacher modeling as intended (and *SDs*)

Item	Overall	T1 (<i>n</i> = 5)	T2 (<i>n</i> = 5)	T3 (<i>n</i> = 4)
Teacher announces the start of modeling	93 (27)	80 (45)	100 (0)	100 (0)
Teacher gives a concrete assignment to students	64 (50)	40 (55)	100 (0)	50 (58)
Teacher uses I-pronoun for think aloud	100 (0)	100 (0)	100 (0)	100 (0)
No teacher-initiated interaction <i>during</i> the think-aloud	67 (50)	50 (58)	50 (58)	100 (0)
Teacher does not simply read the script	58 (51)	25 (50)	75 (50)	75 (50)
Teacher models with focus on text structure	92 (29)	100 (0)	100 (0)	75 (50)
Teacher initiates the reflective discussion as intended	36 (50)	20 (45)	40 (55)	50 (58)
Total	62 (22)	51 (22)	71 (20)	79 (18)

Note. T1: Teacher 1; T2: Teacher 2; T3: Teacher 3

Guided practice

The average implementation fidelity score for guided practice was 58%, but it varied greatly across teachers (Table 5). The manual provided texts and graphic organizers that the teachers could fill out together with their students. The manual contained relatively few directive guidelines for the execution of this lesson phase. This makes sense, as guided practice requires a great deal of finetuning, and teachers have to be responsive to their students. During guided practice, students answered teacher-initiated questions in all observed lessons (100%). Often, different students could react to the same question (83%), sometimes as a way to correct each other’s answers (42%). Teachers typically provided the correct answers (67%), sometimes followed by additional explanation (58%).

TABLE 5
Percentage of lessons with guided practice (GP) as intended (and *SDs*)

Item	Overall	T1 (<i>n</i> = 4)	T2 (<i>n</i> = 5)	T3 (<i>n</i> = 3)
Teacher provides GP as intended by the manual	58 (51)	75 (50)	40 (55)	66 (58)
Students are asked to read the text aloud	67 (49)	75 (50)	80 (45)	33 (58)
Students answer teacher-initiated questions	100 (0)	100 (0)	100 (0)	100 (0)
Several students get to answer the question	83 (39)	75 (50)	80 (45)	100 (0)
Teacher provides correct answers	67 (49)	75 (50)	80 (45)	33 (58)
Students correct each other’s answers	42 (51)	25 (50)	40 (55)	66 (58)
Teacher provides explanation	58 (51)	50 (58)	40 (55)	100 (0)

Note. T1: Teacher 1; T2: Teacher 2; T3: Teacher 3

There were differences in style across teachers. Teacher 1 asked many different questions and often provided the right answers himself, whereas Teacher 3 tended to ask relatively few questions and had students complement each other’s

answers, often followed by a recap of theory. Teacher 2 often provided additional explanation and choral reading, or sometimes omitted activities, as she already added interactive elements during the phase of explicit instruction. Teacher 1 and Teacher 2 expressed their desire for more directions and practice opportunities – in particular for struggling readers – for this lesson phase. By contrast, Teacher 3 felt that there were too many directions for guided practice, and referred in particular to the PowerPoint slides with sample answers for the graphic organizers.

Collaborative learning

In all lessons, teachers provided their students with the opportunity to engage in collaborative learning tasks. The overall implementation score was 50% for the phase of collaborative learning (Table 6), but differed across teachers. Teacher 1 realized a much higher implementation fidelity score (80%) than the other teachers who often deviated from the jigsaw principles in the second phase of the collaborative learning task where students had to exchange information and complete a joint product.

TABLE 6

Percentage of lessons with collaborative learning as intended (and SDs)

Item	Overall	T1 (n = 5)	T2 (n = 5)	T3 (n = 4)
Teacher enacts collaborative task as intended	50 (52)	80 (45)	40 (55)	25 (50)
Teacher gives instruction to promote collaboration	57 (51)	80 (45)	60 (55)	25 (50)
Total	50 (41)	80 (27)	50 (50)	25 (29)

Note. T1: Teacher 1; T2: Teacher 2; T3: Teacher 3

In 50% of the observed collaborative activities, teachers were seen to provide feedback during the process, albeit in different ways. In 80% of the observed lessons, Teacher 1 provided a large amount of spontaneous feedback. In particular, during the second part of the jigsaw task, he provided a high level of

guidance and feedback, stimulating students to exchange information in an efficient way, as explained in (4).

4) They don't know by themselves what to do, which means chaos! So you need to create clear structures and routines. I made sure students were sitting in the right groups, and I was also seated next to a few weaker students. I told them all steps, like: you first tell your classmate what you have learned from the text. Then you listen to what your peers have learned. Then you write down the most important thing that you learned by listening to them. You should mention each step, as small as it may be. (Interview, T1)

Teacher 2 walked around during 60% of the lessons, mainly during the first part of the jigsaw task. She inspected student workbooks, stimulated her students to formulate appropriate responses, and regularly asked students to show to her where in the text they found the information. Teacher 3 never walked around to check students' work and only occasionally answered questions that were raised by his students, but he realized a moment of feedback after task completion.

The teachers identified several implementation challenges related to collaborative learning. Teacher 2 mentioned that it was time consuming to mix students in heterogeneous groups, and that it required a high level of class management skills, although this became less of a problem once the jigsaw became more of a routine. Teacher 1 did not experience class management issues, but mentioned that his students needed a high level of guidance for effective collaboration. He believed that this was a complex skill for his students (5), although he recognized that his students gradually developed better collaborative skills, thanks to the repetition of the jigsaw structure.

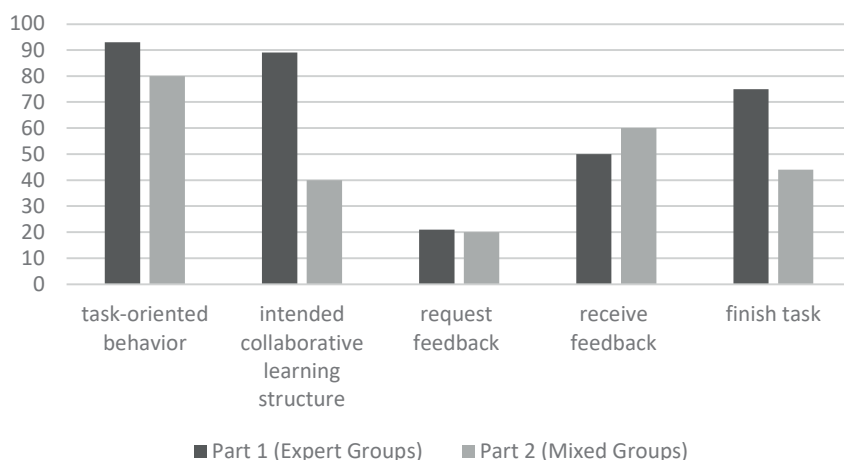
5) Sharing your information in a structured way with your peers, telling what you have learned, and making sure that someone else can learn from you and make notes: that is a very complex skill for children. It requires a lot from them, and from me as a teacher as well! (Interview, T1)

By contrast, Teacher 3 felt that there was too much repetition of the same jigsaw structure, which is why he combined it with a different collaborative learning structure: once expert groups had completed their part of the task and were required to work in mixed groupings on a joint product, the teacher asked students to walk around and exchange information with random classmates. However, this undermined the collaborative nature of the activity: students did not take notes and often compared their answers with students who had read the same text, instead of exchanging information in mixed groups. Furthermore, these exchanges seemed ineffective, as the intended information exchange and joint product were often not achieved.

Figure 3 shows the outcomes of the student observations during the collaborative task. In four of the observed lessons, the collaborative task was slightly different in nature and lacked a mixed group structure, so that only data were collected and analyzed under Part 1. The observational data show that during the first part of the task, students more often worked in the intended collaborative structure (89%) than in the second part where they had to exchange information (40%). During the first part, students showed a higher level of task-oriented behavior (93%) and more often finished the task (75%) than in the second part (80% and 44% respectively).

FIGURE 3

Observed student behavior (%; $n = 28$ pairs) during collaborative learning



Individual practice and lesson closure

The degree of implementation fidelity for the lesson phase of individual practice was 57% (Table 7). Teacher 1 and Teacher 2 both provided feedback during individual practice by walking around, Teacher 3 provided feedback afterwards in two of the observed lessons. As reported in the previous section, the lesson phase of individual practice was often shortened by the teachers, and the same holds true for the lesson closure. That is, the intervention lessons prescribed a lesson closure in which teachers were prompted to reflect with their students on the lesson goals, and discuss an overarching question related to transfer and/or metacognitive knowledge (e.g., *In what other types of texts would you expect a chronological structure? How can you make use of what you learnt today when you read such a chronological text in future?*). Yet, the observed lessons lacked the intended lesson closure: in only 4 out of 14 observed lessons the teachers realized a lesson closure, but this was typically a quick discussion of the individual practice part. As a result, lessons did not contain an explicit take-home message focused on transfer.

TABLE 7
Percentage of lessons with individual practice and closure as intended (and *SDs*)

Item	Overall	T1	T2	T3
		(<i>n</i> = 5)	(<i>n</i> = 5)	(<i>n</i> = 4)
Teacher enacts individual practice as intended	71 (47)	40 (55)	100 (0)	75 (50)
Teacher provides feedback on task as intended	50 (52)	40 (55)	60 (55)	50 (58)
Teacher provides lesson closure as intended	29 (47)	20 (45)	40 (55)	25 (50)
Total	50 (32)	33 (41)	67 (24)	50 (33)

Note. T1: Teacher 1; T2: Teacher 2; T3: Teacher 3

Implementation fidelity, satisfaction, and students’ engagement

The teachers rated the complete lesson series with a mean score of 8.5 out of 10, and were satisfied with the texts and the level of difficulty of the lessons. Table 8 shows that students’ responses on the satisfaction questionnaire are in line with these scores: the level of texts and lesson content was not too easy or difficult, and students reported that the focus on text structure was useful. Students were

on average slightly less positive about the texts ($M = 3.42$) than the teachers ($M = 4.46$), but students showed more variability in their opinions.

Zooming in on teachers' satisfaction with the different phases of the GRR model, it was found that teachers were relatively satisfied with each lesson phase. Even for lesson phases that had a relatively low implementation fidelity score (e.g., modeling, collaborative learning; see previous section), teachers rated these phases as positive. In their logbooks, the teachers made some suggestions on providing more cues for guided practice (Teacher 1), or to raise students' involvement with student whiteboards (Teacher 2), and to narrow the focus of individual practice to the lesson goals by adding a separate transfer lesson (Teacher 3).

TABLE 8

Teacher and student satisfaction scores

Teachers' satisfaction scores ($n = 24$ logbook entries)	Mean (SD)
Overall satisfaction with lesson series (0-10)	8.50 (1.10)
Level of difficulty of lesson content (0-10)	8.50 (0.72)
Appropriateness of texts (0-5)	4.46 (0.72)
Satisfaction: explicit instruction (0-5)	4.58 (0.58)
Satisfaction: modeling and reflection (0-5)	4.58 (0.58)
Satisfaction: guided practice (0-5)	4.14 (1.15)
Satisfaction: collaborative learning (0-5)	4.13 (0.85)
Satisfaction: individual practice (0-5)	4.04 (0.99)
Students' satisfaction scores on a 0-5 scale ($n = 25$ students)	Mean (SD)
I learned a lot about different text structures.	4.42 (0.83)
What I learned about text structures is useful to me.	4.33 (1.17)
I know better how to summarize different texts.	3.88 (0.85)
I found the texts that we read really interesting.	3.42 (1.18)
The texts were too difficult.	2.68 (1.04)
The lessons were too easy for me.	2.83 (1.11)
There was too much repetition in the lessons.	2.67 (1.05)
I liked working in groups.	4.63 (0.58)
I liked to work individually.	3.57 (1.04)
I liked the bigger tasks, like the quiz, poster, or letter.	4.38 (1.28)

The interviews and logbooks revealed teachers' enthusiasm with the collaborative activities: they felt inspired, and reported that students showed more time-on-task behavior and enthusiasm than during regular lessons. Most students of Teacher 2 who completed the questionnaire also highly valued the collaborative learning activities ($M = 4.63$) and the authentic tasks (e.g., letter, quiz) that were provided ($M = 4.38$).

The final phase of the GRR model, individual practice, received slightly lower satisfaction scores from teachers and students. In his logbook, Teacher 1 mentioned that more time could be devoted to individual practice, and that the current list of activities provided for this lesson phase was too limited. Teacher 2 highly valued the focus of the individual practice phase, in particular the repetition of structures that were covered in previous lessons. Teacher 3 reported the opposite and felt that the repetition of structures and the transfer writing tasks were a weakness of the program. He preferred individual practice that was closer to lesson content and felt that the tasks should be less focused on transfer and repetition, in order to provide a better check on the achievement of lesson goals.

Teachers' observed implementation fidelity scores were unrelated to their self-reported satisfaction scores ($r = .02, p = .86$). Only for Teacher 3, these scores were positively correlated ($r = .52, p = .023$), suggesting that he was more satisfied with the lesson phases that were well implemented. Students' involvement scores per lesson phase also appeared unrelated to teachers' implementation fidelity scores ($r = .12, p = .30$). In fact, students' average involvement during the different lesson phases was high ($M = 1.83$ out of 2; $SD = 0.41$), even when lesson phases took longer than intended. However, the previous section discussed how students' observed on-task behavior appeared to be lower during the second part of collaborative learning tasks: the part of the task that was often implemented with lower fidelity scores.

Looking at the appreciation of individual lessons, Lesson 2 on compare-contrast structure received the highest teacher ratings for both overall satisfaction ($M = 9.33$; $SD = 1.15$) and appropriateness for students ($M = 9.00$, $SD = 1.00$). Students also mentioned the lessons on compare-contrast and chronology most often as their favorite lesson ($n = 8$ and $n = 8$, respectively). Some students

explained their answer by stating that the text topic of that lesson was interesting ($n = 11$), that the lesson was easy ($n = 6$), and/or that it had a great task ($n = 2$).

By contrast, Lesson 8 on cause-effect texts received the lowest overall teacher rating ($M = 7.00$; $SD = 0.58$), and was rated relatively low on text quality ($M = 3.67$; $SD = 1.15$) by the teachers. During the interview, teachers explained that the lesson was too difficult in terms of content (i.e., distinguishing direct and indirect effects), text topic (i.e., no prior knowledge on El Niño), and task (i.e., creating an information poster). Not surprisingly, many students ($n = 10$) nominated the cause-effect lessons as their least preferred lessons: these were too difficult ($n = 4$), too long or boring ($n = 4$), and/or not really important for them ($n = 3$). Student workbooks also showed that many students did not finish the task in Lesson 8.

6.4 Conclusion and Discussion

The current study examined teachers' implementation of the GRR model in the context of a text structure intervention (TOP). The findings of this multiple-case study only constitute a starting point for large-scale studies on the quality of reading instruction, as the data concern just three teachers, and the student questionnaire only voices the opinion of the students from one teacher. Because of the limited sample size, it was decided not to establish a link between implementation aspects and student outcomes. Despite these limitations, several findings in this study (e.g., related to difficulties with modeling and facilitating collaborative learning) echo preliminary findings from *Chapter 5* and dovetail with findings from other studies in the Dutch context (Okkinga et al., 2018a; 2021; Veldman et al., 2020). Several findings were confirmed by triangulating data (Brown, 2009; Kuorikoski & Marchionni, 2016; Miles et al., 2013), although it is important to note that the data from teachers logbooks did not always perfectly converge with lesson observations.

The focus of analysis was to examine whether teachers' adaptations in the intervention program would hinder the realization of a successful GRR. Although the GRR model provides researchers with a useful lens to examine teachers' implementation, the GRR model should not be applied with rigidity and undercut teachers' professionalism. Therefore, 100% fidelity of implementation

scores would be both unrealistic and undesirable: in the end, as teaching is an art that requires teachers to flexibly adapt lessons to their students' needs. However, teachers' adaptations during implementation should never undermine the principles of effective instruction (Harn et al., 2013), such as the importance that all students can experience all phases of the GRR model for deep learning (Fisher & Frey, 2021).

With regard to the structural dimensions of implementation fidelity, the data showed that teachers never fundamentally reordered the suggested instruction and activities in a way that could undermine a GRR. Even though missing-the-middle – that is, jumping from explicit instruction to individual practice – is an often reported problem with regard to the realization of a GRR (Dole et al., 2019; Fisher & Frey, 2021), the highly scripted and structured intervention program (TOP) appeared to work well in this regard. That is, the teachers never skipped guided practice and/or collaborative learning, and regularly paid even more time to these activities in the middle than was prompted by the manual, even if this led to time management issues later on.

Zooming in on the process-oriented dimensions of implementation fidelity, there appeared to be considerable differences across the different phases of the GRR. The highest average implementation score was obtained for the explicit instruction phase (91%). For the successive lesson phases, the average implementation scores gradually decreased: modeling (62%), guided practice (58%), collaborative learning (50%), and individual practice and lesson closure (50%). Generally speaking, teachers adhered less to the program during the student-centered activities than during teacher-centered activities. For collaborative learning, this might reflect the fact that this phase is considered to be challenging (Fisher & Frey, 2021); for the final two phases, this can be partially attributed to time constraints at the end of lessons.

There also appeared to be considerable differences across teachers: each teacher realized a specific GRR implementation pattern over lessons, which seemed related to their own preferences and knowledge. Teacher 1 struggled with theoretical concepts during teacher-centered instruction, but was very skillful in guiding collaborative activities. By contrast, Teacher 3 was very skillful in providing clear instruction and modeling – even without relying on the script too much –

but he gave little guidance and feedback during collaborative learning, which challenged the principle of scaffolded instruction that underlies the GRR (Wood et al., 1976). Teacher 2 was very skillful in increasing students' involvement by adding interactive elements to the intervention program (e.g., mini whiteboards, choral reading, interaction during modeling) and by providing high levels of feedback. To some extent, this formed an enrichment to the program: for example, choral reading can improve students' fluency and prosody, and hence contribute to better comprehension (Kuhn, 2020; Turner, 2010), and the use of mini whiteboards can increase students' levels of engagement and on-task behavior (Marsh et al., 2021).

Yet, through the lens of the GRR model, Teacher 2 actually turned the first lesson phases – explicit instruction, modeling, guided practice – all into one extensive form of guided practice, which threatened a GRR where the cognitive load of the learned material should very gradually shift from teacher to student (Fisher & Frey, 2013). The fact that she turned modeling into an interactive discussion also challenges the foundational principles of modeling in which students should assume an observer role in order to avoid cognitive overload while learning complex tasks (as discussed in the context of writing, see Evers-Vermeul & van den Bergh, 2009; Rijlaarsdam et al., 2005).

For two teachers, no positive correlation was found between teachers' self-reported satisfaction with lesson phases and the fidelity of implementation scores they obtained. Apparently, these teachers did not consider their implementation quality as a criterion for their satisfaction, which might suggest that they may hold different criteria to evaluate their implementation, and consequently, may have limited insight into the foundational principles of the GRR. A crucial conclusion is that researchers should always use multiple sources of data and not uniquely rely on teachers' self-reports, as they might use different, or maybe even invalid, criteria for evaluating their implementation quality.

Below, we turn to three main implementation challenges across teachers and lessons that need further reflection. These are related to (1) promoting metacognitive knowledge, (2) conditions for effective collaboration, and (3) a tension between teacher autonomy and the degree of scriptedness.

Metacognitive knowledge

According to Fisher and Frey (2021), students and teachers should engage in metacognitive thinking (i.e., reflecting on their learning and/or task approach) during all phases of the GRR. The TOP intervention explicitly focused on aspects of metacognitive knowledge by providing conditional knowledge goals, prompts for teachers to engage in reflective discussions on their reading approach after they finished modeling, and by providing metacognitive reflection tasks during lesson closure. Instead of simply highlighting the recognition of text structures, we hoped this approach could sensitize students to reasons for when and why certain reading strategies can be applied in a flexible way to promote text comprehension.

Yet, even though teachers briefly mentioned the conditional lesson goal at the start of the lesson, little attention was paid to metacognitive knowledge throughout most lessons. For example, despite the fact that concrete questions were provided in the manual for reflection after modeling (see also Braaksma et al., 2002; Schutz & Rainey, 2019), these were only seldom incorporated by teachers, who often skipped directly to the phase of guided practice. In addition, the intended discussion during lesson closure was also often not realized – which might be partially explained by time management issues (see also Gillies & Boyle, 2010) – so that options for transfer were not explicitly discussed. Yet, teaching for transfer seems crucial for students in this age group (Patton et al., 2022). Even though the individual practice activities also intentionally focused on transfer of text structure knowledge in writing tasks (Hebert et al., 2016; Strong, 2020), one teacher felt that the transfer tasks were a weakness of the program, as it reduced the focus on the exact content of that lesson at hand. This was not reported by the other teachers, but illustrates how aspects related to promoting metacognitive knowledge might be challenging for teachers and/or be at tension with their current level of knowledge.

Collaborative learning

Collaborative learning forms a crucial element in the socio-constructivist inspired GRR model: in small groups, students act as scaffolds to their peers, so that there is a gradual transition from teacher-centered to student-centered learning (Dole

et al., 2019). In fact, practicing reading and understanding texts in small groups can be an effective instructional approach, but a high-quality implementation is critical for success (e.g., Capin et al., 2021; Okkinga et al., 2021).

In the current study, most collaborative learning tasks had a jigsaw structure with expert groups and mixed groups as a way to promote interdependence and joint argumentation (see Aronson, 1978). This use of a variety of groupings is an effective practice to promote learning (Archer & Hughes, 2011), but these different groupings for the jigsaw tasks led – at least in the beginning of the lesson series – to class and time management issues. This is more often reported in studies on collaborative learning (van Kuijk et al., 2021). Fortunately, these practical issues resolved themselves as the lessons progressed, and students and teachers became more familiar with the collaborative learning structure. Yet, it seems recommendable that ample time is scheduled for collaborative learning tasks in the early lessons, so that orderly routines can be established. In addition, teachers might benefit from guidance related to optimal grouping (e.g., group size, group composition), as well as support on how to divide attention among multiple groups (e.g., Capin et al., 2021; Okkinga et al., 2021).

Another implementation issue was the fact that two teachers regularly made unfortunate changes in the collaborative learning structure, mainly during the exchange of information in mixed groups. This undermined the intended positive interdependence of students (Johnson & Johnson, 1987; 2017), and seemed to be reflected in the fact that students displayed less task-oriented behavior during the second part of jigsaw tasks, and finished this part of the collaborative task less often. This illustrates how important and challenging it is for teachers to effectively guide collaborative learning in whole-classroom settings (see Hacker & Tenent, 2002; Okkinga et al., 2021; Spörer et al., 2009), especially with young students who possess few social skills and discourse skills for collaboration (De Backer et al. 2021; Hacker & Tenent, 2002; Veldman, 2020).

Although small groups of students might spontaneously engage in behavior to construct meaning from a text without direct teacher guidance (Pulles et al., 2022), the current study suggests that, without proper teacher guidance, pupil-pupil interactions become of limited value to learning: students got more focused on lower-level aspects or even went off-task, a finding that resonates with previous research as well (Galton et al., 1999; Okkinga et al., 2021).

Teachers' feedback practices seem key to effective implementation: where Teacher 3 provided almost no feedback during collaborative tasks – resulting in very superficial collaborative behavior – and Teacher 2 mainly provided product feedback during the first part of the task, Teacher 1 provided step-by-step guidance and feedback on both the reading task and the type of collaborative behavior. This approach resulted in a high level of student involvement, also during the information exchange in the second part of the jigsaw task. Although difficult at first, Teacher 1 noticed how his students gradually became more able to effectively collaborate and work through the text. This underscores the critical role of primary school teachers in the scaffolding of meaningful dialogues during collaborative learning (see also Baines et al., 2009; Hacker & Tenent, 2002).

Despite the challenges related to collaborative learning, it is an encouraging finding that both students and teachers in the current study highly valued the collaborative learning tasks, which formed quite a contrast with their regular reading curriculum (*Chapter 3*).

Degree of scriptedness

Previous studies suggest that teachers find it difficult to teach about text structures (Beerwinkle et al., 2018; Reutzel et al., 2016). This might also hold true for the Dutch context where teachers receive little instruction about text structures in teacher training colleges (Kooiker-den Boer et al., 2019). Moreover, Dutch teachers typically do not encounter good examples of text structure instruction and/or an effective GRR in their regular curriculum (*Chapter 3*). Therefore, the TOP intervention provided a highly scripted curriculum: each lesson phase – from explicit instruction to lesson closure – was explained step-by-step in a teaching manual, and the highlights of instruction were also shown on the screen. For most lesson phases, the manual provided scripts in the voice of the teacher (e.g., I-pronoun modeling scripts), as this can support a successful implementation (Droop et al., 2016; Schneider, 2013).

Our study shows that the scripted lessons helped teachers to provide text structure instruction without requiring too much lesson preparation. Even though the topic of text structures was quite unfamiliar to the teachers, a high level of implementation fidelity was realized for explicit instruction (91%). Still, the high

degree of teacher guidance also had a downside: modeling scripts and the cues for providing guided practice were sometimes experienced as too tight, or created tension with teachers' own autonomy. This tension seemed to interact with teachers' pedagogical content knowledge. Teacher 3 had a relatively high level of knowledge about modeling and felt constrained by the scripted materials. Teacher 2 initially valued the modeling scripts, but as lessons progressed and her knowledge increased, she reported that the modeling scripts became too prescriptive. Teacher 1 struggled with theoretical concepts and mentioned repeatedly that the teaching materials could be even more scripted: he wanted more scripts, explanation, and examples

Scripted curricula have been praised and criticized by teachers and researchers for many reasons (Demko, 2010; Eisenbach, 2011; Timberlake et al., 2017). Although scripted curricula can support teachers' enactment, the degree of scriptedness can also cause frustration or resistance if there is no sufficient compatibility with teachers' preferences and pedagogical content knowledge (Miller et al., 2017; van Kuijk et al., 2021; Timberlake et al., 2017). Textbooks should support teachers' enactment but not diminish their need to engage in responsive teaching and reflections on the subject matter (Graue et al., 2015).

The finding that teachers' pedagogical content knowledge seemed to interact with their use and appreciation of scripted teaching materials is very relevant in the context of creating (educative) teaching materials (see Ball & Cohen, 1996; Davis & Krajcik, 2005; Davis et al., 2017; Edelson et al., 2021; Haas et al., 2021; Schneider et al., 2005). That is, if teachers' initial knowledge is relatively low, they might benefit from more guidance and scripts; yet, with increasing levels of knowledge, there should be decreasing levels of scaffolding *for teachers* within their manuals.

Implementation research can provide insights into the how and why of intervention success in the messy classroom. It should be conducted not only in the context of effect studies, but also as a way to examine teachers' pedagogical content knowledge. We highly recommend large-scale implementation studies – in particular with a focus on aspects of the GRR model and/or text structure – ideally combined with data on teachers' prior knowledge, and relate findings to student outcomes. Learning from teachers' implementation struggles and

solutions also provides relevant information for creating educative teaching materials that are not too tight but provide support for teachers within their zone of proximal development (Ball & Cohen, 1996; Davis & Krajcik, 2005; Davis et al., 2017; Edelson et al., 2021; Haas et al., 2021; Schneider et al., 2005). For example, the implementation challenges that were identified in this chapter suggest that teachers need more knowledge and support related to effective modeling and promoting student dialogues during collaborative learning.

Instead of providing highly scripted materials throughout the curriculum, it might be wise to gradually proceed from elaborate scripts to open suggestions for modeling and collaboration once teachers get more experienced. At the same time, solutions should not only come from enhanced, evidence-based teaching materials: teachers' implementation struggles may also inspire teacher training colleges to finetune their curriculum (Scheerens & Blömeke, 2016). Dutch teacher training colleges could pay more attention to the ingredients needed for higher-order comprehension and effective reading instruction (van den Broek et al., 2021), for example by discussing text structures (Kooiker-den Boer et al., 2019) and by explaining the principles and implications of the GRR model.

CHAPTER 7



General discussion

7.1 Research aim

An experienced paleontologist who is digging up dinosaur bones instantly generates hypotheses on where and how these bones fit into a skeleton, based on his knowledge about skeletons. Likewise, proficient readers make predictions about text content based on the text's skeleton or underlying text structure. Just like dinosaurs and giraffes have different skeletons, informational texts can have different underlying structures, such as problem-solution, cause-effect, compare-contrast, and chronology (Meyer, 1975). Each of these structures has a characteristic way in which main ideas are organized and connected, which is often expressed through structure-specific signaling words or cue phrases (Armbruster, 2004).

Imagine for instance a book on the worldwide coronavirus pandemic. The first chapter tells about the discovery of the disease and how it spread over the world, and the second chapter details the differences and similarities of the coronavirus and the seasonal flu. Probably, the first chapter highlights the order of events and has a chronological structure that contains many temporal connectives (e.g., *after this*, *next*, *a few days later*), whereas the second chapter focuses on a comparison and will be filled with comparison words and antonyms (e.g., *likewise*, *bigger*, *smaller*). For each text structure, different steps can be followed to reach a coherent summary or to create a visualization. The first chapter might be visualized with a timeline or flowchart, while a Venn diagram can highlight the main ideas from the second chapter. Also, different questions can be asked in each text, and different predictions can be made about upcoming text content for each text structure (Hoch & McNally, 2020; Oakhill et al., 2015; Stevens & Vaughn, 2021).

Readers who do not use the text structure often rely on a default list-strategy to process information (Meyer et al., 1980; Rapp et al., 2007), whereas proficient readers typically attend to text-structural features in order to select the essential text ideas, and analyze how these all fit together (Anderson & Pearson, 1984; Meyer & Rice, 1984), just like the paleontologist did when examining the

excavated bones. Even though not all informational texts are well-structured, and even though texts often display a mixture of different text structures, knowledge of the five basic structures (Meyer, 1975) may be foundational to understanding texts with more complex structures later on (e.g., Meyer et al., 2018).

In order to promote higher-order text comprehension skills, it has been proposed that reading comprehension instruction should pay attention to reading strategies and text structures (Aarnoutse, 2017; Duke et al., 2011; Meyer, 2018; Ogle & Blachowicz, 2002). Therefore, the main focus of this dissertation is whether and how theory on text structures can be translated into practice, that is, in Dutch primary schools. Of course, this is only one possible approach to stimulate reading comprehension, as it is a very complex and multifaceted skill (e.g., Cain et al. 2004; Cromley & Azevedo, 2007; Snow, 2002; Verhoeven & Perfetti, 2008).

The majority of students who participated in this research were fourth or fifth graders. At this age, students are increasingly exposed to informational texts and expected to engage more and more in reading-to-learn behavior (Adams, 2009; Chall & Jacobs, 1983). National and international assessments suggest that primary school students in the Netherlands often struggle with higher-order text comprehension (e.g., van den Broek et al., 2021) and typically engage in superficial text processing when dealing with informational texts (Kim & van Dusen, 1998; PIRLS-2016: Gubbels et al., 2017; PPON-54: Kuhlemeier et al., 2014). Although international studies have already shown that fourth and fifth graders can benefit from text structure instruction (Hebert et al., 2016; Pyle et al., 2017), this topic receives very limited attention in Dutch schools (Scheltinga et al., 2013) and in the teacher training curriculum (Kooiker-den Boer et al., 2019).

The main purpose of this dissertation was therefore to examine the feasibility and effects of text structure instruction in the Dutch context. I also paid attention to issues related to instructional practices, teacher knowledge, and the role of textbooks, as the quality of this so-called ecological component of reading comprehension is crucial for good outcomes (Aaron et al., 2008). This final chapter summarizes and discusses the main findings (7.2), followed by a reflection on important themes and limitations as a guidance for future research (7.3). The last section discusses several practical recommendations (7.4).

7.2 Main findings

The importance of text structure instruction (Chapter 2)

The meta-analysis in *Chapter 2* examined whether explicit text structure instruction has demonstrable immediate and delayed effects for students in the upper elementary grades. The results of 44 high-quality studies were combined and summarized in this meta-analysis. *Chapter 2* revealed that explicit text structure instruction has positive immediate effects on reading outcomes for students in grades 4 to 6. However, at delayed posttests, the differences between text structure groups and control groups were no longer demonstrable. The effect sizes of text structure instruction varied across outcome measures with Hedges' $g = 0.25$ for comprehension questions, $g = 0.37$ on text recall tasks, $g = 0.57$ on summarization, and $g = 0.38$ on text structure knowledge. These effect sizes are realized in authentic classrooms, over and above business-as-usual gains, which makes them meaningful for educational practice.

The effects of text structure instruction varied across outcome measures. The largest effect sizes were found with summarization and recall tasks. Probably, each type of outcome measure taps into a slightly different constellation of comprehension skills (Brown et al., 1983; Hare & Borchardt, 1984). Changes in reading comprehension due to text structure instruction might be better shown with tasks closely related to situation model comprehension skills than with outcome measures that tap into text base comprehension (Kintsch, 1988; 2013). This has implications for deciding what constitutes an adequate measure to evaluate the effects of text structure instruction (Clemens & Fuchs, 2021; see 7.3).

The results from the meta-analysis provide guidance for the design of text structure interventions for the upper elementary grades. For example, the moderator analysis showed that text structure interventions are even more powerful when additional instruction is provided on paragraph-level structures (i.e., making a distinction between topic sentences and details within paragraphs), or on how to create or complete structure-based graphic organizers (e.g., flowcharts, cause-effect schemas). Also, the rule-based summarization strategy, which is a step-wise approach to select main ideas and create a summary, had an additional positive impact on summarization and recall measures.

In order to inspire the quest for an optimal instructional approach, we also analyzed the impact of instructional characteristics of the interventions. This moderator analysis suggested that students might benefit from ample opportunities for individual practice, in order for retention of effects to occur. An additional effect of modeling or collaborative activities could not be demonstrated. However, it is important to note that not all studies provided elaborate descriptions and/or illustrations of instructional approaches, so that the analysis had to be based on the reported instructional approach, without checks on the actual quality of these instructional approaches (see also section 7.3).

In a nutshell, the meta-analysis in Chapter 2 demonstrated positive effects of explicit text structure instruction for students in the upper elementary grades, and most clearly showed an effect on outcome measures that tap into higher-order comprehension skills. I agree with other educators and researchers that text structure instruction should be a key component of the primary school reading curriculum as a way to promote higher-order comprehension skills (e.g., Duke et al., 2011; Hogan et al., 2011; Houtveen & van Steensel, 2022; Shanahan et al., 2010; Wijekumar et al., 2021). Therefore, the main results of the meta-analysis were translated into a set of recommendations in a handbook on effective reading comprehension instruction for Dutch teachers and teacher trainers (Bogaerds-Hazenberg et al., 2022b).

The research-practice gap (Chapter 3)

In order to examine whether a research-practice gap exists, the mixed-methods study in *Chapter 3* analyzed the way in which reading strategies and text structure instruction are taught in the Dutch curriculum for grades 4 and 5. These findings have been confronted with the main conclusions from the meta-analysis and other influential theories on the development of higher-order reading comprehension, stating that students need reading instruction that addresses three types of knowledge – declarative, procedural, conditional knowledge – in a balanced way (Alexander, 2018; Cross & Paris, 1988; Duke & Pearson, 2008; Paris et al., 1983). In addition, the Gradual Release of Responsibility Model (GRR model; Fisher & Frey, 2021; Pearson & Gallagher, 1983) was used to examine the instructional approach as suggested by textbooks and realized by teachers.

The main findings of *Chapter 3* are as follows: (1) textbooks and teachers overemphasize the practice of reading strategies through texts-with-questions, at the expense of attention to conditional knowledge; (2) textbooks and teachers pay limited and fragmented attention to text structure; (3) the suggested and realized instructional approach are not always well-aligned with the GRR model.

First, textbooks and teachers devote much time to students developing procedural knowledge through practicing reading strategies in fixed tasks that sometimes dictate exactly what students need to do next (e.g., *Look at the title and pictures. What do you think this text is about?*), and through answering questions about a text. Questions are often not well aligned with lesson goals, and have to be answered in a vacuum: functional contexts or reading goals are lacking. Almost no attention is paid to the conditional when-and-why knowledge that is needed for a self-regulated, flexible strategy use, which makes it questionable whether this text-with-questions approach contributes to improved reading comprehension skills (see also 7.3). Not surprisingly, several teachers that were interviewed mention issues related to a lack of self-regulation and transfer.

Second, the current curriculum lacks coherent instruction about text structures. Lesson observations and interviews confirm this picture as well: teachers' main focus is on procedural how-to knowledge about strategies. Several teachers mentioned text structure related aspects like summarization, signaling words, and main idea identification among the topics their students struggled with most. Yet, teachers often do not compensate for their textbooks in this regard, possibly due to their own limited knowledge about effective text structure instruction, as has been reported in other research (Beerwinkle et al., 2018; Reutzel et al., 2016; Wijekumar et al., 2019a).

Third, both in classroom practice and in textbooks, a gradual release of responsibility is often not realized in the instructional approach. Great emphasis is put on the lesson phase where students have to answer comprehension questions individually; an activity that on average takes up 42% of the lesson time. This activity is both highly valued and criticized by the teachers. In addition, the implementation of collaborative learning and modeling is not optimal. Collaborative learning hardly goes beyond working parallel in pairs on the same questions, and the modeling scripts that teachers often rely on are of limited quality, as they are mingled with interactive questions and fail to make crucial

(metacognitive) steps in the thinking process explicit. Also, these scripts do not prompt teachers to engage in a reflection on what happened during the think-aloud, even though this is considered a crucial component of effective modeling (Schutz & Rainey, 2019).

These findings do not appear to be very different from the ones in international textbook research: reading curricula often pay limited attention to text structures (Austria: Seifert, 2021; China: Zhang et al., 2021; Croatia: Peti-Stantić et al., 2021; US: Wijekumar et al., 2021) and put emphasis on text content, often at the expense of promoting higher-order thinking strategies (Malta: Aguis & Zammit, 2021; Portugal: Cordeiro et al., 2021; US: Beerwinkle et al., 2021).

Compared to these international textbook studies, *Chapter 3* also added the layer of the enacted curriculum, showing that most teachers follow their textbooks to the letter, and do not compensate for flaws or gaps in the curriculum – even when they diagnose several problems. Possibly, teachers lack pedagogical content knowledge to compensate for the low quality of their books, and to provide high quality instruction on text structure and reading strategies by themselves. This has been reported in other countries as well (e.g., Beerwinkle et al., 2018; Dockx et al., 2020; Wijekumar et al., 2019a).

The main message of *Chapter 2* and *Chapter 3* is therefore that, even though coherent and explicit instruction about text structures is relevant in the primary school curriculum, this type of instruction is currently lacking in the Dutch curriculum. The research-practice gap also extends to the way in which strategic reading skills are taught: there is no balance in declarative, procedural, and conditional knowledge, and the release of responsibility in instructional activities is not as gradual as one might hope for. Hence, the third part of the dissertation (*Chapter 4, 5* and *6*) describes my attempt to reduce the research-practice gap by developing and testing text structure lessons in close collaboration with teachers.

Reducing the research-practice gap (Chapter 4)

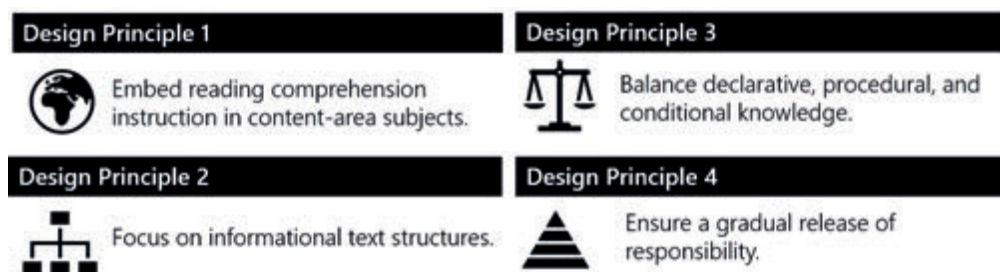
Just like engineers make different decisions about the shape or materials for a bridge – sometimes in a process of trial and error – our bridge between research and practice was designed in iterative design cycles. First, the viability of four evidence-based design guidelines for lessons on text structure was examined, as

well as whether they could be applied by the co-designing teachers (*Chapter 4*). After two design cycles, the effects of these lessons were analyzed in a large-scale intervention study with a main focus on student outcomes (*Chapter 5*). As the experiment raised some questions on implementation fidelity, a multiple-case study was conducted (*Chapter 6*).

Chapter 4 details the construction process of the TOP text structure intervention. Four teachers and two researchers co-designed lessons on text structure, based on design principles (DPs) that were derived from literature and findings from the previous chapters (see Figure 1). During the iterative design and testing of the lessons, various types of data were collected: lesson artefacts and feedback, teacher logbooks, lesson observations, and panel interviews. This approach is typical of design-based research (DBR), a methodology that is often proposed as a way to bridge an existing research-practice gap as it can produce evidence-based materials, and simultaneously provide valuable insight into the viability of underlying design principles, and is likely to promote teacher professionalization as well (Bannan-Ritland, 2008; Broekkamp & van Hout-Wolters, 2007; Edelson et al., 2021; McKenney, 2013).

FIGURE 1

Overview of the design principles for the text structure lessons (*Chapter 4*)



Chapter 4 shows that not all four design principles are easy to implement, which seems related to various factors: (1) text quality, (2) conflicting ideas and priorities, and (3) teachers' limited pedagogical content knowledge.

Text quality was a first bottleneck issue. As content-area subject can offer a motivating, meaningful, and effective context for reading comprehension

instruction (Duke et al., 2021; Hwang et al., 2021; Maerten-Rivera et al., 2016), the co-designing teachers in our study strived to integrate DP1 and DP2 by selecting informational texts from their content-area textbooks. Yet, the text quality turned out to be limited with regard to text structural aspects: texts were often descriptive in nature, or contained a mixture of different text structures.

This finding is in line with other studies, showing that many Dutch content-area texts form a complex mixture of narrative and expository elements (Sangers et al., 2021), and often lack a clear text structure beyond the sentence level (Kooiker-den Boer et al., to appear). In a first attempt, two teachers revised the texts for their lessons, but this was a time-consuming endeavor with dissatisfying results. The two other teachers intentionally used authentic, non-revised texts, but this undermined the focus on text structure instruction: their students were not exposed to clear examples to learn the basic characteristics of each text structure, and much lesson time had to be devoted to discussing other aspects of the text, such as the complex vocabulary. This finding highlights the need for educational publishers to provide more well-structured texts for primary education, so that teachers will have access to well-structured, high-quality texts. Familiarizing students with basic informational text structures (see Meyer, 1975) forms an important foundational stepping stone for students to eventually understand more complex text structures and texts with a less clear structure later on (e.g., Meyer et al., 2018).

The intentional use of authentic, often poorly structured texts by two of the co-designing teachers, is also indicative of the second issue that complicated the lesson design: conflicting ideas and priorities, which is a well-known issue in DBR research (Cviko et al., 2014; McKenney & Reeves, 2018). For example, two teachers felt that explaining vocabulary of authentic texts should always take a central place in lesson design – even if this meant fewer opportunities for text structure instruction. In the DBR study, conflicting ideas on effective modeling approaches also hit the surface: although the value of coping models (i.e., making mistakes and correcting them while thinking aloud) was explained to teachers, they were afraid to set a wrong example, and instead, always acted as perfect mastery models. Some conflicting ideas were gradually transformed with researchers' laborious support and feedback.

Third, it appeared during the DBR that teachers' knowledge about several topics was quite limited, in particular with regard to informational text structures and the concept of conditional knowledge. Even after being trained in these design principles (DP2 and DP3), teachers still found it difficult to recognize text structures themselves, or to model a structure-specific strategy use. They also struggled with formulating conditional lesson goals and developing activities and instruction focused on this when-and-why-knowledge.

In sum, *Chapter 4* shows that the wide research-practice gap presents challenges for short-term DBR projects. The limited attention to text structures and the lack of expertise on conditional knowledge that was uncovered in *Chapter 3* with regard to the current curriculum, also hit the surface in the design-based research. This issue became even more complicated due to the limited quality of the content-area texts that were selected. Although good results can be achieved with a DBR project, it can be challenging to engage teachers as co-designers, because they are not always aware of their limited knowledge and/or different perspectives. A successful DBR requires a considerable time investment from teachers as well as from researchers who need to provide extensive support throughout the process.

The proof of the pudding: effects and implementation (Chapter 5-6)

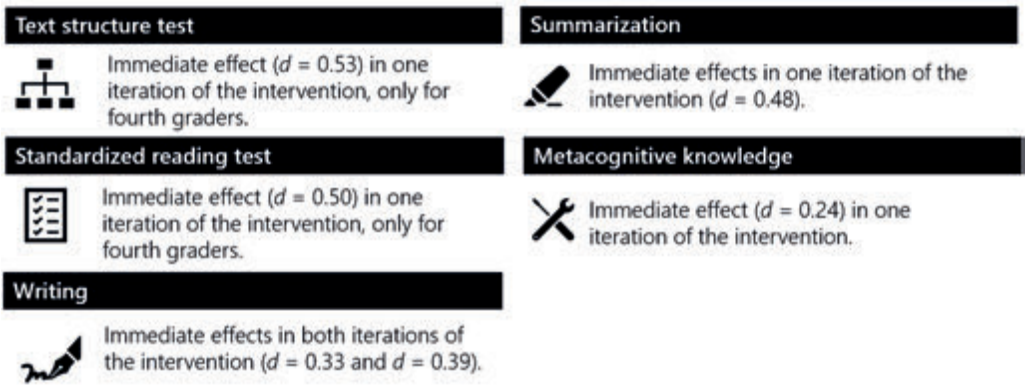
Ten text structure lessons that were developed during the DBR were selected and revised in such a way that each lesson focused on one of four informational text structures: compare-contrast, problem-solution, chronology, and cause-effect. The aim of the resulting TOP intervention was that students would be able to recognize these structures, summarize them in graphic organizers, and learn about a structure-specific application of reading strategies such as questioning.

Chapter 5 evaluated the effects of the TOP intervention. Ten Dutch teachers of grades 4-6 participated with their students ($N = 201$) in an experiment with a switching-panels replication design. They either first followed the TOP intervention and then went back to business-as-usual, or the other way around (Shadish et al., 2002). There were three measurement occasions, so that the immediate effects of the intervention could be estimated in two iterations of the intervention, as well as delayed effects in one group. Effects were measured for

reading comprehension, summarization skills, explicit metacognitive knowledge, and writing. Some data on implementation fidelity were collected as well.

Figure 2 shows the intervention effects per outcome measure. Different from our expectations, only one out of four possible immediate effects could be demonstrated: only fourth graders in one iteration of the intervention showed positive immediate effects on the text structure test ($d = 0.50$) and the standardized reading test ($d = 0.53$). The fourth and fifth graders in this group also showed immediate effects on the summarization task ($d = 0.48$) and the metacognitive knowledge test ($d = 0.24$). An immediate effect of the intervention on students’ writing skills was demonstrable in all groups ($d = 0.33$ and $d = 0.39$). We were unable to show delayed effects.

FIGURE 2
Outcomes of the intervention study (*Chapter 5*)



For the group that did show the expected effects on all outcome measures, the findings corroborate the meta-analytic findings of *Chapter 2*. That is, knowledge about text structure seems to help students identify main ideas, and organize these in a meaningful framework, which facilitates both reading comprehension and summarization (Meyer et al., 1980; Miyatsu et al., 2018; Stevens et al., 2021; Strong, 2020; Taylor, 1985; Westby et al., 2010; Wijekumar et al., 2012). The transfer effect on students’ writing skills lends some support for the theory that knowledge

of text structures not only provides readers with a useful frame to organize incoming information in their mind, but also provides writers with a means to organize their ideas on paper (Dickson, 1999; Hebert et al., 2018). However, as for most outcome measures the intervention groups did not show the same pattern of effects, we cannot generalize this finding over all students.

It is intriguing that only the fourth graders in one group showed positive immediate effects on all outcome measures, while the only replicable effect in the other groups was found for writing. On the one hand, this finding warrants a reflection on the quality of the outcome measures that were used within the restraints that the switching-panels replication design poses on parallel tests (see 7.3). On the other hand, it raises questions about implementation fidelity and intervention robustness. The content of the intervention was quite complex while its duration was relatively short when compared to similar interventions. Some intervention components might have been implemented differently than intended, or the implementation might have varied across groups. For example, the limited implementation data suggested that collaborative learning tasks were not always implemented as intended, and that several teachers regularly skipped individual practice activities. Therefore, a multiple-case study was set up to examine implementation aspects.

The multiple-case study in *Chapter 6* focuses on the ways in which three teachers enacted a slightly revised version of the text structure intervention, and what implementation issues arose during different phases of the GRR model; that is, during explicit instruction, modeling, guided practice, collaborative learning, individual practice, and lesson closure. This focus fits in nicely with the call for second-generation text structure research focusing on the conditions under which interventions are successful (Strong, 2020; Williams, 2018). The data came from logbooks, lesson observations ($N = 14$), teacher interviews, student questionnaires, and observations of students' behavior during collaborative learning activities.

The data show that both teachers and students valued the focus on informational text structures and liked the collaborative jigsaw activities, which form quite a contrast with the dominant text-with-questions approach in Dutch textbooks (*Chapter 3*). With regard to the structural dimension of implementation fidelity, the data showed that teachers never fundamentally reordered the

suggested instruction and activities in a way that could undermine a GRR. The highly scripted intervention program appeared to work well in this regard.

Zooming in on the process-oriented dimensions of implementation, there appeared to be differences across the different phases of the GRR. The highest average implementation score was obtained for explicit instruction (91%). For the successive lesson phases, the average implementation scores were lower and gradually decreased: modeling (62%), guided practice (58%), collaborative learning (50%), and individual practice and lesson closure (50%). Teachers adhered less to the program during the student-centered activities than during teacher-centered activities, but each teacher had a unique style in how they enacted the GRR model.

Of course, teaching is an art that requires teachers to flexibly adapt lessons to their students' needs, but these adaptations should not undermine the principles of effective instruction (Harn et al., 2013). Some implementation issues were observed across teachers and lessons, which are related to (1) promoting metacognitive knowledge, (2) promoting effective collaboration, and (3) a tension between teacher autonomy and limited teacher knowledge.

First, the TOP intervention highlighted not only the recognition of different text structures, but also the way in which knowledge of text structures can contribute to finetuning one's strategy use. In order to promote students' metacognitive knowledge, the intervention suggested reflective discussions after modeling focused on teachers' reading approach, as well as a lesson closure in which teachers and students discussed when and why this newly acquired knowledge could be most useful. Yet, the intended reflective discussions and lesson closures were often skipped, which might have limited students' opportunities to gain metacognitive knowledge.

Second, although students and teachers highly valued the collaborative learning activities, the implementation of collaborative learning appeared difficult. Two teachers struggled to promote an effective information exchange between their students, and all teachers sometimes made changes in the second part of the collaborative learning task in a way that undermined students' positive interdependence, which is a critical condition for collaboration (Johnson &

Johnson, 1984; 2017). This seemed to have a negative impact on student behavior: during the second part of the task, students showed less task-oriented behavior.

This corroborates previous research on collaborative learning: pupil-pupil interactions easily get off-task (Galton et al., 1999). It is challenging for teachers to guide collaborative learning in classroom settings (Hacker & Tenent, 2002; Okkinga et al., 2018b; Spörer et al., 2009) even after several years of training and coaching trajectories (Okkinga et al., 2021). This is especially true in primary school settings, as children's social skills and discourse skills are still relatively limited (De Backer et al. 2021; Veldman et al., 2020). Yet, the success of one of the teachers in our study to engage his students in joint argumentation by providing them with step-by-step guidance shows that, with an effort, good results can be achieved.

Third, the implementation data showed a challenge in providing enough curricular support for teachers without undermining their professional autonomy. As previous studies have shown that teachers find it difficult to teach text structures and apply reading strategies in expository texts (e.g., Beerwinkle et al., 2018; Capin et al., 2021; Kucan et al., 2011; Reutzel et al., 2016), the intervention contained many texts and scripts in the voice of the teacher, as this can support teachers' implementation (e.g., Droop et al., 2016; Schneider, 2013). This approach of a scripted curriculum worked well during the design study (*Chapter 4*), and also seemed to help the teachers in this multiple-case study to provide text structure instruction without requiring too much lesson preparation. Yet, the downside was that teachers sometimes relied too much on these scripts when their knowledge was limited, or that they felt constrained by the scripted materials when they had a high level of knowledge. For example, one of the teachers initially valued the modeling scripts, but as her confidence and knowledge increased, she felt that the scripts became too prescriptive.

These implementation issues resonate findings that already popped up during the lesson design (*Chapter 4*), and to some extent reflect the issues with the current curriculum (*Chapter 3*). For example, the limited attention to true collaborative learning activities and conditional knowledge development in the current curriculum (*Chapter 3*) seems to leave teachers unfamiliar with these concepts. This makes it difficult for teachers to effectively implement collaborative and metacognitive aspects in their lesson design (*Chapter 4*), and also results in difficulties while teaching lessons (*Chapter 6*). The same holds true for modeling:

although comparing models and/or showing a model that hesitates, makes mistakes, and then corrects these, can be valuable for students, especially the weaker ones, this coping model approach is not applied in current teaching materials (*Chapter 3*) and was also criticized by the teachers of the design study (*Chapter 4*), and the multiple-case study (*Chapter 6*).

It is important to learn from teachers' implementation struggles, and to be sensitive to the interaction with their pedagogical content knowledge. Research on implementation fidelity can specify the conditions for intervention success, but can also yield valuable insight into the support and additional knowledge that teachers need for effective enactment. This can inspire the design of teacher-friendly educative teaching materials, which simultaneously address student learning and teachers' own development of pedagogical content knowledge (Ball & Cohen, 1996; Davis & Krajcik, 2005; Davis et al., 2017; Edelson et al., 2021; Haas et al., 2021; Schneider et al., 2005).

7.3 Discussion of main findings

The status of text structure instruction

This dissertation showed the importance of text structure instruction and the need to integrate it in the current Dutch primary school curriculum. In order to prevent that text structure is simply added to existing materials as yet another activity or piece of knowledge, we need to clarify the status of text structure instruction first.

In many reading curricula, text structure receives little attention and is nothing more than a separate activity after students work through a text (Wijekumar et al., 2019b; 2021). By contrast, researchers emphasize that text structure is not an independent activity or piece of knowledge, but should be considered as a powerful reading strategy that supports readers during inference making, main idea generation, summarizing, and so on (Meyer & Wijekumar, 2007; Meyer et al., 2018; Wijekumar et al., 2019b). Therefore, text structure is often listed among other evidence-based reading strategies such as rereading, visualizing, or questioning (e.g., Duke & Pearson, 2008; Janzen & Stoller, 1998; Yang, 2006).

However, if we look closely at text structure research, it seems inaccurate to describe text structure as yet another strategy in readers' repertoire. Rather, text structure knowledge is a vehicle to fine-tune students' use of reading strategies. Therefore, I would argue for a different viewpoint: text structure should be considered as a meaningful framework for contextualized strategy use. That is, instead of having students apply reading strategies as abstract entities, text structure provides them with a concrete tool of how to apply strategies in a contextualized way, that is, fit to the text structure at hand.

This is precisely what most text structure interventions already do: students learn to differentiate between structures and adapt their strategy use accordingly. A compare-contrast text is visualized differently from a problem-solution text, different steps can be followed to reach a coherent summary, and different predictions can be made about upcoming text content (Hoch & McNally, 2020; Stevens & Vaughn, 2021; Wijekumar et al., 2019b). With this approach, the overall conceptualization of reading strategies remains the same, but it is interpreted more directly, with emphasis on text signals (Williams, 2005). This relates to the notion that readers should supplement bottom-up reading strategies with top-down strategies (Stanovich & Stanovich, 1995).

The meta-analysis in *Chapter 2* showed that this finetuning of reading strategies to specific text structures (e.g., structure-based summarization) is what appears to be more powerful than general summarization techniques. This resonates with theories on strategic reading: only if students know how, when, and why to apply reading strategies in a way that fits the text and task at hand, will they become self-regulated, strategic readers (Afflerbach et al., 2008; Alexander, 2018; Cross & Paris, 1988; Nash-Ditzel, 2014; Paris et al., 1983). From this perspective, it is not surprising that the context-blind strategy practice in the current curriculum (*Chapter 3*) does not seem to produce self-regulated readers who transfer their knowledge outside the direct context of reading lessons. That is, once strategies are overemphasized as a goal in itself, this inhibits students' self-regulated, flexible use of strategies (Sinatra et al., 2002; Wilkinson & Son, 2011). Yet, strategy instruction that heavily relies on relatively abstract metacognitive skills might not be suitable for children either (Williams, 1998).

Text structure instruction can provide a solid middle ground: teaching reading strategies within the meaningful framework of text structure can make

these strategies concrete yet generalizable enough for students to be applied by themselves outside the direct context of reading lessons. It is promising that in all studies where text structure lessons were taught (*Chapters 4-6*), the teachers mentioned – often to their surprise – instances of spontaneous transfer. In a nutshell, text structure should not be another strategy or isolated piece of knowledge, but should be integrated as a concrete framework for strategy instruction.

Towards a powerful instructional approach: the value of the GRR model

In order to generate insight into effective components of reading comprehension instruction, research should focus not only on high-quality content (*what* we teach), but also on examining powerful instructional approaches (*how* we teach) (Pyle et al., 2017; Williams, 2018). Throughout this dissertation, the study, design, and discussion of instructional approaches was centered around the influential GRR model (Fisher & Frey, 2021; Pearson & Gallagher, 1983). This section reflects on the usefulness of this model for research, curriculum design, and practice.

The GRR model: its value for research

The GRR model appears to be useful as framework for analytical purposes, as it resulted in good interrater reliability scores in both the meta-analysis (*Chapter 2*) and the textbook analysis (*Chapter 3*). Yet, two threats to using the GRR model were identified: underspecification and misclassification.

First, the problem of underspecification: most research papers in the meta-analysis did not provide sample materials and/or elaborate descriptions of the intended instructional approach, so that the meta-analysis was based on the *reported* instructional approach, without direct quality checks on instructional aspects. The fact that we could not prove an additional effect of modeling and/or collaborative learning might be due to a limited implementation quality (e.g., Okkinga et al., 2018b; 2021)

Second, misclassification impeded the analysis of the highly scripted textbooks that were analyzed (*Chapter 3*): several activities were labeled by textbook publishers as modeling, whereas a close inspection revealed that modeling was often no true think aloud, but rather a teacher-led interactive

discussion with students, often focused on text content and much less on the reading approach (see Schutz & Rainey, 2019). Similarly, activities labeled as collaborative learning often turned out to be nothing else than answering questions individually but parallel in pairs, which ignores the five conditions for effective collaboration (Johnson & Johnson, 1984; 2017).

These issues with underspecification and misclassification underscore that more work needs to be done in order to move the research forward. A crucial first step is to provide elaborate and systematic descriptions of instructional approaches, for example by using the terminology offered in the GRR model, and by providing details on the design principles of interventions, at both the macro-level (i.e., focus and mode of instruction, sequencing of content) and the micro-level (e.g., instructional activities, learning activities, materials) (Bouwer & De Smedt, 2018). The next step would be to systematically test hypotheses derived from the GRR model on the value of different instructional features such as individualized feedback or web-based tutoring, as exemplified in text structure studies by Meyer et al. (2002, 2010) and Wijekumar et al. (2012, 2013, 2014).

The GRR model as a design guideline

The GRR model was a useful design guideline for co-designing teachers who used it as a framework for sequencing instruction and activities in the lesson series on text structure (*Chapter 4*). Still, some questions were raised. For example, establishing a full gradual release in each single lesson made lessons too long. Some lessons were shortened and revised for the intervention study (*Chapter 5*), but the full GRR model was kept in place, as the program was too short to divide it over multiple lessons. Various teachers reported that they shortened their lessons by altering the collaborative activities or skipping individual activities. For the multiple-case study (*Chapter 6*), several texts were shortened, whereas the design of collaborative and individual activities became more structured, so that it could not be skipped that easily. Even though minor time management issues remained, the execution of the instructional approach was better. Two teachers mentioned how application of the GRR in their lessons was useful and set clear expectations for their students.

These findings show that the GRR guidelines are useful for lesson design, although it seems important to reflect on ways to implement the GRR over lessons

in a flexible way. It remains a balancing act between offering a structured instruction, while simultaneously providing room for teachers' flexibility to fit instruction to students' needs and practical demands (Webb et al., 2019).

The GRR model: its value for classroom practice

The value of the GRR model for classroom practice is in the first place related to its directions for the sequencing of activities: there should be a *gradual* – not an abrupt – release of responsibility: independent practice ('I do it') only follows after adequate support from teachers ('We do it together') and from peers ('You do it together') (Fisher & Frey 2021; Pearson & Gallagher, 1983). The current Dutch curriculum has a limited GRR due to an overemphasis on individual activities, few opportunities for collaborative learning, and a limited quality of teacher-centered activities. This seems partially due to practical reasons and the testing culture that makes teachers prone to teaching to the test (*Chapter 3*; see Cheng & Curtis, 2004; Clemens & Fuchs, 2021).

International research has also found indications that teachers tend to offer independent practice too quickly in reading comprehension lessons, and often fail to provide adequate scaffolding through guided practice and collaboration (Almasi & Fullerton, 2012; Hudson, 2021; Shanahan et al., 2019), which is referred to as the missing-the-middle problem (Dole et al., 2019; Fisher & Frey, 2021). A more consistent implementation of the GRR model in the Dutch context would imply to stop overemphasizing individual practice, and to strengthen the quality of teacher-centered instruction and collaborative learning.

The value of the GRR model is not only related to its clear directions for optimal instructional sequencing, but also in the fact that it provides indicators for successful implementation of each subcomponent. For example, if modeling is stressed as a teacher-centered activity *before* students gain responsibility, it suggests that modeling should not be undertaken as an interactive classroom activity in which students act as participants instead of observers, as has been argued for writing instruction (compare Evers-Vermeul & van den Bergh, 2009; Rijlaarsdam et al. 2005). Likewise, if collaborative activities are implemented as a way to scaffold students through interaction with their peers, it means that

collaborative activities should enable students to act as support for their peers; not to work parallel in pairs without positive interdependence.

These issues with modeling and collaborative learning were identified in the current curriculum (*Chapter 3*) and also came up during the following studies. During the DBR study (*Chapter 4*), great emphasis was put on a high-quality implementation of modeling and collaborative learning, and the co-designing teachers received intensive guidance; yet, once teachers were guided less intensively while implementing the lessons (*Chapter 5* and *Chapter 6*), the implementation of modeling and collaborative learning appeared challenging. This only underscores the high investment and skillfulness it requires from teachers (see also Hacker & Tenent, 2002). Moreover, the fact that for most teachers in the multiple-case study, their satisfaction with lesson phases appeared unrelated to the implementation scores they obtained, suggests that teachers may have limited insight in GRR derived criteria for a successful approach.

Of course, an instructional model such as the GRR should not be considered as an unchangeable rule. Although it provides useful directions, teaching is and will always be a balancing act in which teachers have to respect curricular guidelines and pedagogical principles, as well as adapt these in a flexible way that fits their students' needs. Yet, teachers' adaptations should facilitate, instead of undermine the intended learning process. Hence, it is important to examine how teaching materials can provide teachers with scaffolding cues for each GRR phase, without creating implementation rigidity (McVee et al., 2018; Shanahan et al., 2019). We need to find ways to support teachers in the particular aspects of modeling and collaborative learning that created challenges (e.g., metacognitive aspects, acting as coping models, ensuring positive interdependence in collaborative tasks), as described in this dissertation (*Chapters 4-6*), as well as in other Dutch research (Okkinga, 2018b; 2021; Veldman et al., 2020).

Chapter 2 suggested that individual practice is an important part of the GRR model, but the quality of individual practice in the current Dutch curriculum leaves room for improvement: many questions do not target higher-order comprehension skills, and/or are not properly aligned with lesson goals and instruction, and typically lack a clear context of a functional task (*Chapter 3*). This is quite disturbing, as instructional models emphasize that, when teaching complex tasks, all part-task practice should be provided within a functional task

in order to facilitate transfer (4C/ID-model; van Merriënboer & Kirschner, 2013). In particular, fourth and fifth graders seem to benefit from explicit teaching for transfer in the area of expository text comprehension (Patton et al., 2022).

The texts-with-questions approach might not improve students' higher-order comprehension skills and transfer of these skills, but instead create superficial readers who heavily rely on questions to comprehend a text (Rooijackers et al., 2021; van de Wetering & Groenendijk, 2015). It should be a priority of educational publishers and schools to provide a curriculum with constructive alignment between learning goals, instruction, and activities (Rouffet et al., 2022; Scheerens, 2017), and to offer reading tasks that are focused on higher-order comprehension skills in functional contexts (e.g., reading-to-write or reading-to-learn). The studies in this dissertation are encouraging, as many teachers are willing to change their practice.

The teacher and the textbook

The studies in this dissertation revealed various issues related to the quality of textbooks and teachers, which is concerning, as both textbooks and teachers play a crucial role in the quality of instruction. This raises questions about the way in which these issues with textbooks and teachers should be addressed. Some have argued that the key to improvement is the optimization of curricular materials, while others have argued that the key is to invest in teachers' knowledge. Based on the studies in this dissertation, I would argue that we need to acknowledge the mutually dependent, two-way relationship between textbooks and teachers, as teachers moderate the impact of textbooks by their pedagogical content knowledge, while textbooks affect teacher knowledge, by shaping what they should or could do (Ball & Cohen, 1996; Brown, 2009; Davis, 2021; Harwood, 2017; Scheerens & Blömeke, 2016; Valencia et al., 2006). Therefore, we should simultaneously address the optimization of teaching materials and the professionalization of teachers – instead of optimizing only one of them separately.

This perspective creates room for textbooks and other instructional materials to become a vehicle for teachers' professional learning, as they present a vision of what an approach to teaching and learning looks like in practice, and support

teachers in bringing that vision to life in the classroom (Davis, 2021; Edelson et al., 2021). In the DBR study (*Chapter 4*), the co-design of instructional materials served as a site for professional learning, although this approach turned out to be challenging and time consuming for both teachers and researchers. A more feasible approach might be to design textbooks as so-called *educative curricular materials*. Such teaching materials are designed with the intention of supporting both teacher and student learning, for example by clarifying the curriculum and underlying design principles, and explaining instructional choices (Ball & Cohen, 1996; Edelson et al., 2021; Haas et al., 2021; Schneider et al., 2005).

My recommendation would be that researchers and educational publishers explore whether high-quality educative curricular materials can be developed for the teaching of reading comprehension, as a vehicle to broaden and deepen teachers' pedagogical content knowledge, and/or to create accompanying professionalization modules for existing materials to support teachers' enactment (e.g., Wijekumar et al., in preparation). This way, teacher professionalization will become embedded within optimized teaching materials: a true reform by the book. This might not only be one of the least time-consuming ways to improve current practice, at least for the teachers involved, but might also be a very powerful way, as teachers will directly apply their newly acquired knowledge in practice; one of the critical components of successful professional development programs (see Didion et al., 2020; Hudson et al., 2021; Penuel et al., 2007).

This approach deviates from the opinion that teachers should move away from scripted textbooks, and transform from curriculum mediators into curriculum makers who actively design their own lessons (Schnabel et al., 2016). In fact, more and more schools want to let go of their textbooks for reading comprehension (Oosterloo, 2010). This reaction to the reading crisis is understandable as *Chapter 3* showed how teachers heavily relied on their textbooks to the point that these textbooks blinded them for the actual curriculum. This is unfortunately a well-known problem with scripted curricula (Dewitz & Jones, 2013; Graue et al., 2015; Timberlake et al., 2017; Valencia et al., 2006) that can negatively affect the quality of classroom instruction and student outcomes (e.g., Piasta et al., 2009).

Yet, a one-way investment in teachers while throwing out textbooks is problematic, as the quality of the curriculum will become highly dependent on the knowledge of individual teachers, while textbooks guarantee to some extent

a standardized curriculum. *Chapter 4* illustrated that curriculum design that involves experienced teachers is not all sunshine and roses: even with intensive supervision and evidence-based input from researchers, lesson design is difficult and time consuming. Teachers who design lessons themselves are likely to copy some of the pitfalls of the curriculum they are familiar with, so that they might end up with old wine in new bottles. Other countries experimenting with teachers as curriculum designers also report similar problems: teachers are often unable to specify the critical content and miss a clear curricular structure to rely on (Hughes & Lewis, 2020).

A one-way investment in a high-quality curriculum is not sufficient in itself either. In the research reported on in this dissertation, I highly invested in creating evidence-based teaching materials, but relatively little attention was paid to teacher professionalization, which was restricted to organizing a single training session (*Chapter 5*) or creating four crash course videoclips (*Chapter 6*) focused on the theory behind the lesson series. Not surprisingly, the outcomes of the intervention study and multiple-case study revealed implementation issues related to teachers' yet limited knowledge about text structure, conditional knowledge, and features of effective collaborative learning and modeling. These issues were not resolved by merely offering a scripted curriculum, which shows again how curriculum improvement is a two-way endeavor: investing in teaching materials must go hand in hand with investing in teachers.

Assessments of reading comprehension

The complex nature of reading comprehension processes presents researchers with a challenge when choosing or designing assessments to evaluate intervention effects. A common approach is to avoid putting all eggs in one basket, by making strategic use of multiple outcome measures that tap into different processes related to text comprehension (Bohaty et al., 2015; Calet et al., 2019; Clemens & Fuchs, 2021; Hebert et al., 2016; Leslie & Caldwell, 2009; Nation & Snowling, 2011). This is why the text structure intervention was evaluated with multiple outcome measures (*Chapter 5*), representing near to far transfer (Clemens & Fuchs, 2021). The expected intervention effects were demonstrated

on all outcome measures, but unfortunately, only for the fourth graders in one iteration of the intervention.

To some extent, this finding should be interpreted by taking into account the issue with test parallelism: the switching-panels replication design requires three perfectly parallel constructed tests for each measurement occasion (i.e., tests with equal means, observed and true score variances) so that changes in scores between measurement occasions can unequivocally be interpreted as growth. All reading tests had to be developed from scratch, and all tests needed to be perfectly parallel; yet, unfortunately, the tests were not (see *Chapter 5*). Therefore, the data analysis was focused on interaction effects, instead of growth patterns, which makes it more difficult to demonstrate intervention effects. The construction of reading tests – in particular with a focus on text structure and higher-order comprehension – should be an important focus of future research, so that questions on growth, maintenance, and replicability can be answered.

The inconsistent findings also warrant a reflection on the degree of assessment alignment of each outcome measure with the intervention content (Clemens & Fuchs, 2021). For example, standardized reading comprehension tests are often used to formally evaluate and compare interventions (e.g., Cain & Oakhill, 2006; Slavin et al., 2009), but they do not necessarily target higher-order, situation model comprehension and/or text structure knowledge (Hebert et al., 2016; Leslie & Caldwell, 2009; van den Broek & Kremer, 2000), which is why the standardized test should be considered as a far transfer test (Clemens & Fuchs, 2021; Hebert et al. 2016). This could explain the inconsistent effects, at least in part. Similarly, the metacognitive knowledge test (Brand-Gruwel, 1995) had a low degree of alignment, as it did not measure a structure-specific, contextualized strategy use but focused on non-contextualized strategy knowledge.

The inconsistent effect on the text structure test can be partially explained by issues with test parallelism and validity. In hindsight, instead of a well-aligned, direct, near transfer measure of intervention effects, the text structure test turned out to be a mid-transfer test at best (Clemens & Fuchs, 2021). That is, several graphic organizer items and texts were inspired by a study among secondary school students (Land, 2009) and did not perfectly resemble the graphic organizers of the intervention. Moreover, several items did not measure students' text structure knowledge directly, but required multiple steps of reasoning. For

example, instead of asking students which cue words can be found in cause-effect texts, students were asked: ‘Peter reads a text on the consequences of unemployment; what signaling words will he probably encounter in the text?’ This requires students to infer in a first step that this text probably has a cause-effect structure, and in a second step, recall which signaling words are characteristic of cause-effect structures. This had not been practiced during the intervention.

Several important lessons can be drawn. First, future reading tests should be provided with systematic descriptions of the underlying components that are targeted, as it helps to decide which tests are suitable for intervention evaluation (Cain & Oakhill, 2006; Keenan et al., 2008). Second, researchers should not uniquely focus on standardized comprehension tests as the gold standard: it remains crucial to develop well-aligned, high-quality, near-transfer measures as well, as these can produce a more nuanced and meaningful evaluation of interventions (Clemens & Fuchs, 2021). In particular, it seems relevant to develop valid and reliable text structure tests, as such tests are not currently available for the Dutch context. For example, a future text structure test might be improved by replacing some items with cloze tasks to measure students’ signaling word knowledge, and by adding multiple-choice questions on main idea identification and inference making based on the text structure (e.g., Wijekumar et al., 2017).

Of course, the inconsistent outcomes of the effect study should be considered from a broader perspective in which implementation issues played a role. The reflection on assessment should therefore not only focus on *what* types of tests we used, but also on the question *when* we should start administering tests to measure effects. First, students completed the post-test measures after ten text structure lessons, which might have been too soon. Although the meta-analysis in *Chapter 2* did not show a moderating effect of intervention duration, and other meta-analyses suggested that it is probably a complex interplay of the intervention duration with the actual number of text structures that matters (e.g., Hebert et al., 2016; Pyle et al., 2017), it is important to reflect on the TOP intervention. That is, all instruction on four different text structures was reduced to ten lessons, which might have been challenging for teachers and students. It might have been wiser to spread the intervention content over more lessons, and to move the moment of testing forward. Second, it would have been better to

conduct a multiple-case study focused on implementation aspects *before* testing an intervention on a large scale, so that some adaptations could have been made in the teaching materials and/or teacher training module.

7.4 Building bridges between research and practice: three recommendations

All in all, this dissertation has shown the value of investing in higher-order reading comprehension skills through explicit text structure instruction, which can serve as a meaningful framework for strategy instruction. Yet, most studies in this dissertation also showed that the research-practice gap is wide. Several recommendations need follow-up in order to reduce this gap.

First, instead of offering isolated strategy practice (procedural knowledge), the curriculum should also enable students to discover why and when certain reading strategies work best (conditional knowledge) and explain how the application of strategies differs across text structures (declarative knowledge). If reading strategy instruction becomes intertwined with text structure instruction, this will create more depth and coherence in the curriculum, as it shifts the focus towards higher-order comprehension and reading-to-learn skills. Moreover, it offers opportunities to develop reading comprehension skills beyond the borders of the narrow reading comprehension curriculum in a way that does justice to both specific knowledge-building lesson goals and generalizable reading-related goals (see also Duke et al., 2021; Williams et al., 2014). A critical condition for this to happen in the Dutch context is that educational publishers start providing well-structured texts for all content-area subjects (Kooiker-den Boer et al., to appear), so that students can experience themselves how knowledge about reading strategies and text structures is useful when preparing for a test, writing task, or oral presentation. This move towards *compreaction* (i.e., *comprehension* leading to *action*; see Duke et al., 2021) is assumed to promote transfer, knowledge-building, and reading motivation.

Second, the instructional approach needs to be reformed. Following the GRR model, the teacher-centered activities should be strengthened with high quality instruction and modeling, improved tasks for true collaborative learning, and less individual practice. Although I acknowledge the importance of constructive alignment (e.g., Scheerens, 2017), textbooks and teachers mimic too often what is

asked on the test. This seems an unfortunate approach to promote higher-order comprehension skills and might have narrowed a considerable part of the curriculum to a collection of texts with comprehension questions that are too often not well aligned with instruction and lesson goals. Not surprisingly, many teachers feel lost in the reading curriculum, which – ironically – makes them even more dependent on their books.

Therefore, my third recommendation is to invest in teachers: their pedagogical content knowledge on reading comprehension should be a first priority in teacher education. For example, future teachers need more knowledge about text structures and how this relates to a contextualized strategy use, as well as more insight into the GRR model and into the features of high-quality modeling and collaborative learning. This requires a simultaneous investment on behalf of teacher training colleges who need to refine their curriculum, and educational publishers who need to improve their textbooks. An important research avenue to professionalize teachers on the job is to explore whether curricular materials for reading comprehension can be transformed into so-called *educative* curricular materials that support both teacher learning and student learning, and develop other types of teacher support in this area as well (e.g., web-based supports, professional development courses).

This dissertation calls for an improvement of the current primary school curriculum for reading comprehension. The research-practice gap can be reduced through a sustainable collaboration between many stakeholders, such as researchers, teachers, (educational) publishers, text writers, test makers, policy makers, and teacher training colleges. A perfect bridge between research and practice is not easily built, but let us at least attempt to create stepping stones between theory and practice, and to gradually reduce the gap, one day at a time.

References & Appendix



References

References marked with an asterisk were included in the meta-analysis (*Chapter 2*).

- Aarnoutse, C.A.J. (1990). Heeft onderwijs in leesvaardigheden of -strategieën effect [Does instruction in reading skills or strategies have an effect]? In M. Boekaerts & E. de Corte (Eds.), *ORD Proceedings* (pp. 173-185). ITS.
- Aarnoutse, C.A.J. (1991). Begrijpend lezen in het basisonderwijs [Reading comprehension in primary education]. In P. Reitsma & M. Walraven (Eds.), *Instructie in begrijpend lezen* (pp. 128-142). Eburon.
- Aarnoutse, C.A.J. (2017). Onderwijs in begrijpend lezen [Reading comprehension instruction]. *Orthopedagogiek: Onderzoek en Praktijk*, 56(11-12), 269-291.
- Aarnoutse, C.A.J., & Schellings, G. (2003). Een onderzoek naar de stimulering van leesstrategieën en leesmotivatie in probleemgestuurde leeromgevingen [A study on promoting reading strategies and reading motivation in inquiry learning]. *Pedagogische Studiën*, 80(1), 110-126.
<https://pedagogischestudien.nl/search?issueld=v80i2>
- Aaron, P.G., Joshi, R.M., Gooden, R., & Bentum, K.E. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities*, 41(1), 67-84.
<https://doi.org/10.1177/0022219407310838>
- Adams, M. (2009). The challenge of advanced texts: The interdependence of reading and learning. In E.H. Hiebert (Ed.), *Reading more, reading better: Are American students reading enough of the right stuff* (pp. 163-189)? Guilford.
- Afflerbach, P., & Cho, B.Y. (2009). Determining and describing reading strategies. In: H.S. Walters & W. Schneider (Eds.), *Metacognition, strategy use, and instruction*, (pp. 201-225). Guilford.
- Afflerbach, P., Pearson, P.D., & Paris, S.G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61(5), 364-373.
<https://doi.org/10.1598/RT.61.5.1>
- Aguis, R., & Zammit, C. (2021). A text structure analysis of fourth and sixth grade Maltese textbooks. *Technology, Knowledge and Learning*, 26(1), 407-428.
<https://doi.org/10.1007/s10758-021-09491-9>
- Alexander, P.A. (2018). Looking down the road: Future directions for research on depth and regulation of strategic processing. *British Journal of Educational Psychology*, 88(1), 152-166. <https://doi.org/10.1111/bjep.12204>
- Almasi, J.F., & Fullerton, S.K. (2012). *Teaching strategic processes in reading*. Guilford.
- * Alvermann, D.E., & Boothby, P.R. (1984). Knowledge of text structure and its influence on a transfer task. *Proceedings of the 68th Annual Meeting of the American Educational Research Association* (pp. 2-20).

- * Alvermann, D.E., & Boothby, P.R. (1986). Children's transfer of graphic organizer instruction. *Reading Psychology*, 7(2), 87-100.
<https://doi.org/10.1080/0270271860070203>
- Anderson, R.C., & Pearson, P.D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P.D. Pearson, R. Barr, M.L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 225-291). Longman.
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), 16-25.
<https://doi.org/10.3102/0013189X11428813>
- Andreassen, R., & Bråten, I. (2011). Implementation and effects of explicit reading comprehension instruction in fifth-grade classrooms. *Learning and Instruction*, 21(4), 520-537. <https://doi.org/10.1016/j.learninstruc.2010.08.003>
- Applegate, A.J., & Applegate, M.D. (2004). The Peter effect: Reading habits and attitudes of preservice teachers. *The Reading Teacher*, 57(6), 554-563.
<https://www.jstor.org/stable/20205399>
- Archer, A., & Hughes, C.A. (2011). *Explicit instruction: Efficient and effective teaching*. Guilford.
- Armbruster, B.B. (2004). Considerate texts. In D. Lapp, J. Flood, & N. Farnan (Eds.), *Content-area reading and learning: Instructional strategies* (pp. 47-58). Erlbaum.
- Armbruster, B.B. & Anderson, T.H. (1982). *Idea-mapping: the technique and its use in the classroom or simulating the 'ups' and 'downs' of reading comprehension* (Reading Education Report No. 36). University of Illinois.
- * Armbruster, B.B., Anderson, T.H., & Meyer, J.L. (1991). Improving content-area reading using instructional graphics. *Reading Research Quarterly*, 26(4), 393-416.
<https://doi.org/10.2307/747895>
- * Armbruster, B.B., Anderson, T.H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, 22(3), 331-346. <https://doi.org/10.2307/747972>
- Aronson, E. (1978). *The jigsaw classroom*. Sage Publishing.
- Baines, E., Rubie-Davies, C., & Blatchford, P. (2009). Improving pupil group work interaction and dialogue in primary classrooms: results from a year-long intervention study. *Cambridge Journal of Education*, 39(1), 95-117.
<https://doi.org/10.1080/03057640802701960>
- Ball, D.L., & Cohen, D.K. (1996). Reform by the book: what is – or might be – the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6-14. <https://doi.org/10.3102/0013189X025009006>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bannan-Ritland, B. (2008). Teacher design research: An emerging paradigm for teachers' professional development. In A. E. Kelly, R.A. Lesh, & J.Y. Baek (Eds.),

Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching (pp. 246-262). Routledge.

- Barab, S. (2014). Design-based research: A methodological toolkit for engineering change. In R.K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 151-170). Cambridge University Press.
- Barron, R.F. (1980, December). *A systematic research procedure, organizers, and overviews: An historical perspective*. Paper presented at the 30th Annual Meeting of the National Reading Conference, San Diego, CA.
- Bartels, M., Rietveld, M.J.H., van Baal, G.C.M., & Boomsma, D. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research*, 5(6), 544-552. <https://pubmed.ncbi.nlm.nih.gov/12573186/>
- Bartlett, B.J. (1978). *Top-level structure as an organizational strategy for recall of classroom text* [Unpublished doctoral dissertation]. Arizona State University.
- Baumann, J.F. (1983). Children's ability to comprehend main ideas in content textbooks. *Literacy Research and Instruction*, 22(4), 322-331. <https://doi.org/10.1080/19388078309557722>
- * Bean, T.W., & Steenwyk, F.L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior*, 16(4), 297-306. <https://doi.org/10.1080/10862968409547523>
- Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506-521. <https://doi.org/10.1037/0022-0663.74.4.506>
- Bednarz, N., Desgagné, S., Maheux, J.F., & Zajc, L.S. (2012). La mise au jour d'un contrat réflexif comme régulateur de démarches de recherche participative: Le cas d'une recherche-action et d'une recherche collaborative [A reflexive contract to regulate the process of participatory research: The case of action research and a collaborative study]. *Recherche en Éducation*, 14(1), 128-151. <https://doi.org/10.4000/ree.5894>
- Beerwinkle, A.L., Owens, J., & Hudson, A. (2021). An analysis of comprehension strategies and skills covered within Grade 3-5 reading textbooks in the United States. *Technology, Knowledge and Learning*, 26(1), 311-338. <https://doi.org/10.1007/s10758-020-09484-0>
- Beerwinkle, A.L., Wijekumar, K., Walpole, S., & Aguis, R. (2018). An analysis of the ecological components within a text structure intervention. *i*(9), 1-24. <https://doi.org/10.1007/s11145-018-9870-5>
- Bell, B., & Gilbert, J. (2005). *Teacher development: A model from science education*. Routledge.
- Berends, R. (2011). Begrijpend leesonderwijs: Zin en onzin [Reading comprehension instruction: Sense and nonsense]. *Tijdschrift Taal voor Opleiders en Onderwijsadviseurs*, 2(3), 22-29.

- * Berkowitz, S.J. (1986). Effects of instruction in text organization on sixth-grade students' memory for expository reading. *Reading Research Quarterly*, 21(2), 161-178. <https://doi.org/10.2307/747843>
- Biancarosa, G., & Snow, C.E. (2004). *Reading next: vision for action and research in middle and high school literacy*. Alliance for Excellent Education.
- Bimmel, P. (2001). Effects of reading strategy instruction in secondary education: A review of intervention studies. *L1 Educational Studies in Language and Literature*, 1(3), 273-298. <https://doi.org/10.1023/A:1013860727487>
- Bimmel, P., & Oostdam, R. (1999). Strategietraining en leesvaardigheid; Theorie en praktijk in balans. *Levende Talen Magazine*, 86(543), 556-564. <https://lt-tijdschriften.nl/ojs/index.php/lrm/article/view/973>
- Bissonnette S., Richard, M., Gauthier, C., & Bouchard, C. (2010). Quelles sont les stratégies d'apprentissages fondamentaux auprès des élèves en difficulté de niveau élémentaire? Résultats d'une méga-analyse [What are fundamental learning skills for struggling primary school students? Results from a mega-analysis]. *Revue de recherche appliquée sur l'apprentissage*, 3(1), 1-34. <https://r-libre.telug.ca/776/>
- Block, C.C., & Duffy, G.G. (2008). Research on teaching comprehension. In C.C. Block & S.R. Paris (Eds.), *Comprehension instruction: Research-based best practices* (Volume 2, pp. 19-37). Guilford.
- Blömeke, S., Hsieh, F.J., Kaiser, G., & Schmidt, W.H. (2014). Knowledge at the end of teacher education: International perspectives. In W.H. Schmidt, F.J. Hsieh, G. Kaiser, & S. Blömeke (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn* (pp. 61-208). Springer.
- Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J., & van den Bergh, H.H. (2017). Inhoud en didactiek van begrijpend lezen [Content and pedagogy of reading comprehension instruction]. *Tijdschrift Taal voor Opleiders en Onderwijsadviseurs*, 8(12), 21-30.
- Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J., & van den Bergh, H.H. (2019). Teachers and researchers as co-designers? A design-based research on reading comprehension instruction in primary education. *Journal of Educational Design Research*, 3(1), 1-23. <https://doi.org/10.15460/eder.3.1.1399>
- Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J., & van den Bergh, H.H. (2021). A meta-analysis on the effects of text structure instruction on reading comprehension in the upper elementary grades. *Reading Research Quarterly*, 56(3), 435-462. <https://doi.org/10.1002/rrq.311>
- Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J. & van den Bergh, H.H. (2022a). What textbooks offer and what teachers teach: An analysis of the Dutch reading comprehension curriculum. *Reading and Writing*, 35, 1497-1523. <https://doi.org/10.1007/s11145-021-10244-4>

- Bogaerds-Hazenberg, S.T.M., van den Bergh, H.H., & Evers-Vermeul, J. (2022b). Onderwijs in tekststructuur [Text structure instruction]. In T. Houtveen & R. van Steensel (Eds.), *De zeven pijlers van onderwijs in begrijpend lezen* (pp. 59-83). Stichting Lezen.
- *Bohaty, J.J. (2015). *The effects of expository text structure instruction on the reading outcomes of 4th and 5th graders experiencing reading difficulties* [Doctoral dissertation, University of Nebraska-Lincoln]. <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1230&context=cehdsdiss>
- Bohaty, J.J., Hebert, M.A., Nelson, J.R., & Brown, J.A. (2015). Methodological status and trends in expository text structure instruction efficacy research. *Reading Horizons*, 54(2), 35-65. https://scholarworks.wmich.edu/reading_horizons/vol54/iss2/3
- *Boothby, P.R., & Alvermann, D.E. (1984). A classroom training study: The effects of graphic organizer instruction on fourth graders' comprehension. *Reading World*, 23(4), 325-339. <https://doi.org/10.1080/19388078409557783>
- Bonset, H., & Hoogeveen, M. (2009). *Lezen in het basisonderwijs: Een inventarisatie van empirisch onderzoek naar begrijpend lezen, leesbevordering en fictie* [Reading in primary education: An inventory of empirical research into reading comprehension, reading promotion, and fiction]. SLO.
- Borenstein, M., Hedges, L.V., Higgins, J.P., & Rothstein, H.R. (2011). *Introduction to meta-analysis*. John Wiley.
- Borenstein, M., Higgins, J., Hedges, L., & Rothstein, H.R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8(1), 5-18. <https://doi.org/10.1002/jrsm.1230>
- Bouwer, R., & De Smedt, F. (2018). Considerations and recommendations for reporting instructional writing interventions in journal articles [Editorial]. *Journal of Writing Research*, 10(2), 115-137. <https://doi.org/10.17239/jowr-2018.10.02.01>
- Bouwer, R., & Koster, M.P. (2016). *Bringing writing research into the classroom: The effectiveness of Tekster, a newly developed writing program for elementary students* [Doctoral dissertation, Utrecht University]. https://issuu.com/tekster6/docs/tekster_proefschrift-digi2
- *Bowman, M., & Gambrell, L. (1981, April). *The effects of story structure questioning upon reading comprehension*. Paper presented at the 54th Annual Meeting of the American Educational Research Association, Los Angeles.
- Braaksma, M., Rijlaarsdam, G., & van den Bergh, H.H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology*, 94(2), 405-415. <https://doi.org/10.1037/0022-0663.94.2.405>
- Brasseur-Hock, I.F., Hock, M.F., Kieffer, M.J., Biancarosa, G., Deshler, D.D. (2011). Adolescent struggling readers in urban schools: Results of a latent class analysis.

- Learning and Individual Differences*, 21(4), 438-452.
<https://doi.org/10.1016/j.lindif.2011.01.008>
- Bråten, I., & Anmarkrud, Ø. (2013). Does naturally occurring comprehension strategies instruction make a difference when students read expository text? *Journal of Research in Reading*, 36(1), 42-57. <https://doi.org/10.1111/j.1467-9817.2011.01489.x>
- *Braxton, D.M. (2009). *The effects of two summarization strategies using expository text on the reading comprehension and summary writing of fourth- and fifth-grade students in an urban, Title 1 school* [Doctoral dissertation, University of Maryland]. <https://drum.lib.umd.edu/handle/1903/9918>
- Britt, M.A., Rouet, J.-F., & Durik, A.M. (2018). *Literacy beyond text comprehension: A theory of purposeful learning*. Routledge.
- Broekkamp, H., & van Hout-Wolters, B. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation*, 13(3), 203-220.
<https://doi.org/10.1080/13803610701626127>
- *Broer, N.A., Aarnoutse, C.A.J., Kieviet, F.K., & van Leeuwe, J.F.J. (2002). The effect of instructing the structural aspect of texts. *Educational Studies*, 28(3), 213-238.
<https://doi.org/10.1080/0305569022000003681>
- Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), 141-178. https://doi.org/10.1207/s15327809jls0202_2
- Brown, A.L., Day, J.D., & Jones, R.S. (1983). Macro-rules for summarizing texts: The development of plans for summarizing texts. *Child Development*, 54(4), 968-979.
<https://doi.org/10.2307/1129901>
- Brown, M.W. (2009). The teacher-tool relationship: Theorizing the design and use of curriculum materials. In J.T. Remilliard, B.A. Herbel-Eisenmann, & G.M. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17-36). Routledge.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88(1), 18-37.
<https://doi.org/10.1037/0022-0663.88.1.18>
- Bruggink, M., & Netten, A.R. (2017). The Netherlands: Education policy and curriculum in reading. In I.V.S. Mullis, M.O. Martin, S. Goh, & C.S. Prendergast (Eds.), *PIRLS-2016 Encyclopedia* (pp. 1-11). TIMSS and PIRLS International Study Centre. <http://pirls2016.org/wp-content/uploads/encyclopedia-pirls/downloadcenter/3.%20Country%20Chapters/Netherlands.pdf>
- Burke, M.D., Boon, R.T., Hatton, H., & Bowman-Perrott, L. (2015). Reading interventions for middle and secondary students with emotional and behavioral

- disorders: A quantitative review of single-case studies. *Behavior Modification*, 39(1), 43-68. <https://doi.org/10.1177/0145445514547958>
- Bus, A.G., & van IJzendoorn, M.H. (1999). Phonological awareness in early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91(3), 403-414. <https://doi.org/10.1037/0022-0663.91.3.403>
- Cain, K. (1999). Ways of reading: How knowledge and use of strategies are related to reading comprehension. *British Journal of Developmental Psychology*, 17(2), 293-309. <https://doi.org/10.1348/026151099165285>
- Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76(4), 697-708. <https://doi.org/10.1348/000709905X69807>
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96(1), 31-42. <https://doi.org/10.1037/0022-0663.96.1.31>
- Capin, P., Stevens, E.A., Stewart, A.A., Swanson, E., & Vaughn, S. (2021). Examining vocabulary, reading comprehension, and content knowledge instruction during fourth grade social studies teaching. *Reading and Writing*, 34(1), 1143-1170. <https://doi.org/10.1007/s11145-020-10106-5>
- *Carriedo, N., & Alonso-Tapia, J. (1996). Main idea comprehension: Training teachers and effects on students. *Journal of Research in Reading*, 19(2), 128-153. <https://doi.org/10.1111/j.1467-9817.1996.tb00093.x>
- Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, 31(2), 199-218. <https://doi.org/10.1177/1098214010366173>
- Chall, J.S., & Jacobs, V.A. (1983). Writing and reading in the elementary grades: Developmental trends among low SES children. *Language Arts*, 60(5), 617-626. <https://doi.org/41961511>
- Chambliss, M., & Calfee, R. (1998). *Textbooks for learning: Nurturing children's minds*. Blackwell Publishing.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing* (pp. 25-40). Routledge.
- Cheng, R. (2019). *Effects of teaching text structure in science text reading: A study among Chinese middle school students* [Doctoral dissertation, Columbia University]. <https://doi.org/10.7916/d8-3pgb-3q94>
- Ciullo, S., & Reutebuch, C. (2013). Computer-based graphic organizers for students with LD: A systematic review of literature. *Learning Disabilities Research & Practice*, 28(4), 196-210. <https://doi.org/10.1111/lrdp.12017>

- Clark, S.K., Helfrich, S.R., & Hatch, L. (2017). Examining preservice teacher content and pedagogical content knowledge needed to teach reading in elementary school. *Journal of Research in Reading*, 40(3), 219-232. <https://doi.org/10.1111/1467-9817.12057>
- Clemens, N.H., & Fuchs, D. (2021). Commercially developed tests of reading comprehension: gold standard or fool's gold? *Reading Research Quarterly*, 57(2), 385-397. <https://doi.org/10.1002/rrq.415>
- Cohen, E.G., & Benton, J. (1988). Making groupwork work. *American Educator*, 12(3), 10-17.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Collins, C., & Pressley, M. (2002). *Comprehension instruction: Research-based best practices*. Guilford.
- Collins, N.D. (1994). Metacognition and reading to learn. *ERIC Digest*, ED376427. <https://files.eric.ed.gov/fulltext/ED376427.pdf>
- Colosi, J.C., & Zales, C. (1998). Jigsaw cooperative learning improves biology lab course. *Bioscience*, 48(2), 118-124. <https://doi.org/10.2307/1313137>
- Committee Meijerink (2008). *Referentiekader taal en rekenen: De referentieniveaus* [Framework of reference for language and arithmetic: The reference levels]. Expertgroep doorlopende leerlijn taal en rekenen. <https://www.slo.nl/thema/vakspecifieke-thema/nederlands/referentiekader-taal/@4230/drempels-taal/>
- Common Core Standards Initiative (2018). *Common Core State Standards for English Language Arts*. www.corestandards.org/ELA-literacy
- Connor, C., Spencer, M., Day, S.L., Guliani, S., Ingebrand, S.W., McLean, L., & Morrison, F. (2014). Capturing the complexity: Content, type, and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes. *Journal of Educational Psychology*, 106(3), 762-778. <https://doi.org/10.1037/a0035921>
- Cooper, H. (1998). *Synthesizing research*. Sage Publishers.
- Cordeiro, C., Nunes, A., Castro, S.L., & Limpo, T. (2021). Teaching reading comprehension in Portuguese primary and middle schools. *Technology, Knowledge and Learning*, 26(1), 339-354. <https://doi.org/10.1007/s10758-020-09486-y>
- Coté, N., Goldman, S.R., & Saul, E.U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, 25(1), 1-53. <https://doi.org/10.1080/01638539809545019>
- Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on learning and transfer. *Learning and Instruction*, 9(2), 109-142. [https://doi.org/10.1016/S0959-4752\(98\)00040-1](https://doi.org/10.1016/S0959-4752(98)00040-1)

- Cromley, J.G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311-325. <https://doi.org/10.1037/0022-0663.99.2.311>
- Cross, D.R., & Paris, S.G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 80*(2), 131-142. <https://doi.org/10.1037/0022-0663.80.2.131>
- Cumming G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- Cutting, L.E., & Scarborough, H.S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*(3), 277-299. https://doi.org/10.1207/s1532799xssr1003_5
- Cviko, A., McKenney, S., & Voogt, J. (2014). Teachers as co-designers of technology-rich learning activities for early literacy. *Technology, Pedagogy, and Education, 24*(4), 443-459. <https://doi.org/10.1080/1475939X.2014.953197>
- *Darch, C.B., Carnine, D.W., & Kameenui, E.J. (1986). The role of graphic organizers and social structure in content-area instruction. *Journal of Reading Behavior, 18*(4), 275-295. <https://doi.org/10.1080/10862968609547576>
- Davis, E.A., & Krajcik, J.S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3-14. <https://doi.org/10.3102/0013189X0340030>
- Davis, E.A., Palincsar, A.S., Smith, P.S., Arias, A.M., & Kademian, S.M. (2017). Educative curriculum materials: Uptake, impact, and implications for research and design. *Educational Researcher, 46*(6), 293-304. <https://doi.org/10.3102/0013189X17727502>
- De Backer, L., Van Keer, H., & Valcke, M. (2021). The functions of shared metacognitive regulation and their differential relation with collaborative learners' understanding of the learning content. *Learning and Instruction, 77*(1), 1-11. <https://doi.org/10.1016/j.learninstruc.2021.101527>
- *de Jou, G.I., & Sperb, T.M. (2009). Lectura comprensiva: Un estudio de intervención [Comprehensive reading: An intervention study]. *Revista Interamerican Journal de Psicología, 43*(1), 12-21.
- Demko, M. (2010). Teachers become zombies: The ugly side of scripted reading curriculum. *Voices from the Middle, 17*(3), 62-64. <http://www.ncte.org/journals/vm/issues/v17-3>
- Dewitz, P., & Jones, J. (2013). Using basal readers: From dutiful fidelity to intelligent decision making. *The Reading Teacher, 66*(5), 391-400. <https://doi.org/10.1002/TRTR.01134>
- Diakidoy, I.A.N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change.

- Contemporary Educational Psychology*, 28(3), 335-356.
[https://doi.org/10.1016/S0361-476X\(02\)00039-5](https://doi.org/10.1016/S0361-476X(02)00039-5)
- Dickson, S. (1999). Integrating reading and writing to teach compare-contrast text structure: A research-based methodology. *Reading and Writing Quarterly*, 15(1), 49-79. <https://doi.org/10.1080/105735699278297>
- Didion, L., Toste, J.R., & Filderman, M.J. (2020). Teacher professional development and student reading achievement: A meta-analytic review of the effects. *Journal of Research on Educational Effectiveness*, 13(1), 29-66.
<https://doi.org/10.1080/19345747.2019.1670884>
- Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3(3), 231-264.
<https://doi.org/10.1007/s11409-008-9029-x>
- Dockx, J., Bellens, K., & De Fraine, B. (2020). Do textbooks matter for reading comprehension? A study in Flemish primary education. *Frontiers of Psychology*, 10(1), 29-59. <https://doi.org/10.3389/fpsyg.2019.02959>
- Dole, J.A., Duffy, G.G., & Pearson, P.D. (2019). Reflections on the Gradual Release of Responsibility Model: Where we've been and where we're going. In M.B. McVee, E. Ortlieb, J.S. Reichenberg, & P.D. Pearson (Eds.), *The Gradual Release of Responsibility in literacy research and practice* (pp. 245-263). Emerald Publishing.
- Donker, A.S., de Boer, H., Kostons, D., van Ewijk, C.C.D., & van der Werf, M.P.C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11(1), 1-26.
<https://doi.org/10.1016/j.edurev.2013.11.002>
- Doymus, K. (2008). Teaching chemical equilibrium with the jigsaw technique. *Research in Science Education*, 38(2), 249-260. <https://doi.org/10.1007/s11165-007-9047-8>
- Droop, M., van Elsäcker, W., Voeten, M.J., & Verhoeven, L. (2016). Long-term effects of strategic reading instruction in the intermediate elementary grades. *Journal of Research on Educational Effectiveness*, 9(1), 77-102.
<https://doi.org/10.1080/19345747.2015.1065528>
- Duffy, G.G. (2002). The case for direct explanation of strategies. In C. Collins & M. Pressley (Eds.), *Comprehension instruction: research-based best practices* (pp. 28-42). Guilford.
- Duffy, G.G., Roehler, L., Meloth, M., & Vavrus, L. (1986). Conceptualizing instructional explanation. *Teaching and Teacher Education*, 2(3), 197-214.
[https://doi.org/10.1016/S0742-051X\(86\)80002-6](https://doi.org/10.1016/S0742-051X(86)80002-6)
- Duke, N.K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, 35(2), 202-224.
<https://doi.org/10.1598/RRQ.35.2.1>

- Duke, N.K., Pearson, P.D., Strachan, S.L., & Billma, A.K. (2011). Essential elements of fostering and teaching reading comprehension. In S.J. Samuels & A.E. Farstrup (Eds.), *What research has to say about reading instruction* (4th edition, pp. 54-93). International Reading Association.
- Duke, N.K., Ward, A.E., & Pearson, P.D. (2021). The science of reading comprehension instruction. *The Reading Teacher*, 74(6), 662-672. <https://doi.org/10.1002/trtr.1993>
- Durkin, D. (1978). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 14(4), 481-533. <https://doi.org/10.1598/RRQ.14.4.2>
- Durlak, J.A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, 34(9), 917-928. <https://doi.org/10.1093/jpepsy/jsp004>
- Durlak, J.A., & DuPre, E.P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes, and the factors affecting implementation. *American Journal of Community Psychology*, 41(1), 327-350. <https://doi.org/10.1007/s10464-008-9165-0>
- Dutch Inspectorate of Education (2019). *Dyslexieverklaringen: Verschillen tussen scholen nader bekeken* [Dyslexia diagnoses: Examination of differences between schools]. <https://www.onderwijsinspectie.nl/documenten/themarapporten/2019/04/10/dyslexieverklaringen-verschillen-tussen-scholen-nader-bekeken>
- Dutch Inspectorate of Education (2020). *Peil.taal en rekenen: Einde basisonderwijs* [Level.language and arithmetic: End of primary education]. <https://www.onderwijsinspectie.nl/documenten/themarapporten/2020/04/22/peil.taal-en-rekenen-2018-2019>
- Dutch Inspectorate of Education (2022). *Peil.leesvaardigheid: Einde (speciaal) basisonderwijs* [Level.reading: End of special and general primary education]. <https://www.onderwijsinspectie.nl/onderwerpen/peil-onderwijs/documenten/themarapporten/2022/12/13/peil-leesvaardigheid-einde-basisonderwijs-en-speciaal-basisonderwijs-2020-2021>
- Edelson, D.C., Reiser, B.J., McNeill, K.L., Mohan, A., Novak, M., Mohan, L., Affolter, R., McGill, T.A.W., Bracey, Z.E., Deutch Noll, J., Kowalski, S.M., Novak, D., Lo, A.S., Landel, C., Krumm, A., Penuel, W.R., Van Horne, K., González-Howard, M., & Suárez, E. (2021). Developing research-based instructional materials to support large-scale transformation of science teaching and learning: The approach of the OpenSciEd middle school program. *Journal of Science Teacher Education*, 32(7), 780-804. <https://doi.org/10.1080/1046560X.2021.1877457>
- Eisenbach, B.B. (2012). Teacher belief and practice in a scripted curriculum. *Journal of Educational Strategies, Issues, and Ideas*, 85(4), 153-156. <https://www.jstor.org/stable/23212874>
- Eklund, K., Torppa, M., Aro, M., Leppänen, P.H.T., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and

8. *Journal of Educational Psychology*, 107(1), 126-140.
<https://doi.org/10.1037/a0037121>
- Eliata, N.M.A., & Miftakh, F. (2021). Emphasizing text structure strategy instruction to scaffold 11th grade students' reading comprehension. *Journal of Languages and Language Teaching*, 9(4), 411-421. <https://doi.org/10.33394/jollt.vvii.3998>
- ELINET (2016). *European declaration of the right to literacy*. <https://elinet.pro/policy-statements/>
- *Elledge, D.H. (2013). *Improving reading comprehension through explicit summarization instruction* [Doctoral dissertation, University of Cincinnati].
https://etd.ohiolink.edu/!etd.send_file?accession=ucin1367928451&disposition=inline
- Eloranta, A., Närhi, V., Ahonen, T., & Aro, T. (2019) Does childhood reading disability or its continuance into adulthood underlie problems in adult-age psychosocial wellbeing? A follow-up study. *Scientific Studies of Reading*, 23(4), 273-286.
<https://doi.org/10.1080/10888438.2018.1561698>
- Englert, C.S., & Tarrant, K.L. (1993) Educational innovations: Achieving curricular change through collaboration. *Education and Treatment of Children*, 16(4), 441-474. <https://www.jstor.org/stable/42899329>
- *Ermis, S. (2008). Using graphic organizers to facilitate elementary students' comprehension of informational text. In M.M. Foote, F. Falk-Ross, S. Szabo, & M.B. Sampson (Eds.), *Navigating the literacy waters: Research, praxis, and advocacy* (pp. 87-102). College Reading Association.
- Evers-Vermeul, J., & van den Bergh, H.H. (2009). Schrijf voor de lezer: Over effecten van lezersgericht (her)schrijven op de kwaliteit van instructieve teksten [Write for the reader: On the effects of reader-focused (re)writing on the quality of instruction texts]. *Levende Talen Tijdschrift*, 10(2), 14-23.
<https://lt-tijdschriften.nl/ojs/index.php/ltt/article/view/169>
- Feskens, R.C.W., Kuhlemeier, H., & Limpens, G. (2016). *Resultaten PISA-2015 in vogelvlucht* [Results PISA-2015: a bird's eye view]. Cito.
- Fidalgo, R., Harris, K., & Braaksma, M. (2018). *Design principles for teaching effective writing* (Volume 34). Brill Publishers.
- Fisher, D., & Frey, N. (2021). *Better learning through structured teaching: A framework for the gradual release of responsibility* (3rd edition). ASCD.
- *Fitzgerald, J., & Spiegel, D.L. (1983). Enhancing children's reading comprehension through instruction in narrative structure. *Journal of Literacy Research*, 15(2), 1-17.
<https://doi.org/10.1080/10862968309547480>
- Fritz, C.O., Morris, P.E., & Richler, J.J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology*, 141(1), 2-18.
<https://doi.org/10.1037/a0024338>

- Fuchs, D., Cho, E., Toste, J.R., Fuchs, L.S., Gilbert, J.K., McMaster, K.L., Svenson, E., & Thompson, A. (2021). A quasi-experimental evaluation of two versions of first-grade PALS: One with and one without repeated reading. *Exceptional Children*, 87(2), 141-162. <https://doi.org/10.1177/0014402920921828>
- Fuchs, D., Fuchs, L.S., Mathes, P.G., & Simmons, D.C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34(1), 174-206. <https://doi.org/10.3102/00028312034001174>
- Gajria, M., Jitendra, A.K., Sood, S., & Sacks, G. (2007). Improving comprehension of expository text in students with LD: A research synthesis. *Journal of Learning Disabilities*, 40(3), 210-225. <https://doi.org/10.1177/00222194070400030301>
- Galton, M.J., Hargreaves, L., Comber, C., Wall, D., & Pell, A. (1999). *Inside the primary classroom: 20 years on*. Routledge.
- Garbe, C., de Krosse, H., Lafontaine, D., Netten, A., Shiel, G., Strating, H., & Valtin, R. (2016). *ELINET: Literacy in The Netherlands. Country Report: Children and adolescents*. University of Cologne. http://www.elinet.eu/fileadmin/ELINET/Redaktion/user_upload/Netherlands_Long_Report.pdf
- Gardill, M.C., & Jitendra, A.K. (1999). Advanced story map instruction: Effects on the reading comprehension of students with learning disabilities. *The Journal of Special Education*, 33(1), 2-17. <https://doi.org/10.1177/002246699903300101>
- Garibaldi, A.M. (1979). Affective contributions of cooperative and group goal structures. *Journal of Educational Psychology*, 71(6), 788-794. <https://doi.org/10.1037/0022-0663.71.6.788>
- Garner, R. (1987). Strategies for reading and studying expository text. *Educational Psychologist*, 22(3-4), 299-312. <https://doi.org/10.1080/00461520.1987.9653054>
- Gentner, D., & Namy, L.L. (1999). Comparison in the development of categories. *Cognitive Development*, 14(4), 487-513. [https://doi.org/10.1016/S0885-2014\(99\)00016-7](https://doi.org/10.1016/S0885-2014(99)00016-7)
- *Gentry, L.J. (2006). *Comparison of the effects of training in expository text structure through annotation text marking and training in vocabulary development on reading comprehension of students going into fourth grade* [Doctoral dissertation, University of South Florida]. <https://digital.lib.usf.edu/SFS0025739/00001>
- Gersten, R., Fuchs, L.S., Williams, J.P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, 71(2), 279-320. <https://doi.org/10.3102/00346543071002279>
- Gillies, R.M., & Boyle, M. (2011). Teachers' reflections of cooperative learning: A two-year follow-up. *Teaching Education*, 22(1), 63-78. <https://doi.org/10.1080/10476210.2010.538045>

- Goldman, S.R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23(3), 357-398.
<https://doi.org/10.1080/01638539709544997>
- Goldman, S.R., & Rakestraw, J.A. (2000). Structural aspects of constructing meaning from text. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Volume 3, pp. 311-335). Lawrence Erlbaum.
- Gordon, C. (1990). Changes in readers' and writers' metacognitive knowledge: Some observations. *Reading Research and Instruction*, 30(1), 1-14.
<https://doi.org/10.1080/19388079009558029>
- *Gordon, C.J., & Pearson, P.D. (1983). *The effects of instruction in meta-comprehension and inferencing on children's comprehension abilities* (Technical Report No. 277). National Institute of Education.
- Gough, P.B., & Tunmer, W.E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6-10.
<https://doi.org/10.1177/074193258600700104>
- Graesser, A.C. (2007). An introduction to strategic reading comprehension. In D. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 3-26). Lawrence Erlbaum. .
- Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
<https://doi.org/10.1037/0033-295X.101.3.371>
- Graue, M.E., Whyte, K.L., & Karabon, A.E. (2015). The power of improvisational teaching. *Teaching and Teacher Education*, 48, 13-21.
<https://doi.org/10.1016/j.tate.2015.01.014>
- *Griffin, C.C., Malone, L.D., & Kameenui, E.J. (1995). Effects of graphic organizer instruction on fifth-grade students. *The Journal of Educational Research*, 89(2), 98-107. <https://doi.org/10.1080/00220671.1995.9941200>
- Gubbels, J., Netten, A., & Verhoeven, L. (2017). *Vijftien jaar leesprestaties in Nederland: PIRLS-2016* [Fifteen years of reading achievement in The Netherlands: PIRLS 2016]. Expertisecentrum Nederlands, Radboud University.
- Gubbels, J., van Langen, A.M.L., Maassen, N.A.M., & Meelissen, M.R.M. (2019). *Resultaten PISA-2018 in vogelvlucht* [Results PISA-2018: a bird's eye view]. University of Twente.
- Gudmundsdottir, S., & Shulman, L. (1987). Pedagogical content knowledge in social studies. *Scandinavian Journal of Educational Research*, 31(2), 59-70.
<https://doi.org/10.1080/0031383870310201>
- Haas, A., Januszyk, R., Grapin, S.E., Goggins, M., Llosa, L., & Lee, O. (2021). Developing instructional materials aligned to the next generation science standards for all students, including English learners. *Journal of Science Teacher Education*, 32(7), 735-756. <https://doi.org/10.1080/1046560X.2020.1827190>

- Hacker, D.J. & Tenent, A. (2002). Implementing reciprocal teaching in the classroom: Overcoming obstacles and making modifications. *Journal of Educational Psychology*, 94(4), 699-718. <https://doi.org/10.1037/0022-0663.94.4.699>
- Hall, K.M., Sabey, B.L., & McClellan, M. (2005). Expository text comprehension: Helping primary-grade teachers use expository texts to full advantage. *Reading Psychology*, 26(3), 211-234. <https://doi.org/10.1080/02702710590962550>
- Hamp-Lyons, L. (1997). Washback, impact, and validity: Ethical concerns. *Language Testing*, 14(3), 295-303. <https://doi.org/10.1177/026553229701400306>
- Hansen, C.L. (1978). Story retelling used with average and learning disabled readers as a measure of reading comprehension. *Learning Disability Quarterly*, 1(3), 62-69. <https://doi.org/10.2307/1510938>
- Hare, V.C., & Borchardt, K.M. (1984). Direct instruction of summarization skills. *Reading Research Quarterly*, 20(1), 62-78. <https://doi.org/10.2307/747652>
- Haria, P.D., & Midgett, E. (2014). A genre-specific reading comprehension strategy to enhance struggling fifth-grade readers' ability to critically analyze argumentative text. *Reading and Writing Quarterly*, 30(4), 297-327. <https://doi.org/10.1080/10573569.2013.818908>
- Harlaar, N., Dale, P.S., & Plomin, R. (2007). From learning to read to reading to learn: Substantial and stable genetic influence. *Child Development*, 78(1), 116-131. <https://doi.org/10.1111/j.1467-8624.2007.00988.x>
- Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children*, 79(2), 181-193. <https://doi.org/10.1177/001440291307900204>
- Harwood, N. (2017). What can we learn from mainstream education textbook research? *RELC Journal*, 48(2), 264-277. <https://doi.org/10.1177/0033688216645472>
- Hebert, M., Bohaty, J.J., Nelson, J.R., & Brown, J. (2016). The effects of text structure instruction on expository reading comprehension: A meta-analysis. *Journal of Educational Psychology*, 108(5), 609-629. <https://doi.org/10.1037/edu0000082>
- Hebert, M., Bohaty, J.J., Nelson, J.R., & Roehling, J.V. (2018). Writing informational text using provided information and text structures: An intervention for upper elementary struggling writers. *Reading and Writing*, 31(9), 2165-2190. <https://doi.org/10.1007/s11145-018-9841-x>
- Hedges, L.V., & Hedberg, E.C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29(1), 60-87. <https://doi.org/10.3102/0162373707299706>
- Heesters, K., van Berkel, S., van der Schoot, F., & Hemker, B. (2007). *Balans van het leesonderwijs aan het einde van de basisschool: uitkomsten van de vierde peiling in 2005. PPON-reeks 33*. [Balance of reading instruction at the end of primary education: findings of the 4th assessment in 2005. PPON series 33]. Cito.

- Heidelberg, J., van den Bergh, H.H., Kuhlemeier, H., & Hoeks, J. (1993). Tijdbesteding van docenten in het voortgezet onderwijs op verschillende manieren gemeten: Een validatiestudie [Different measures of teachers' time allocation in secondary education: A validation study]. *Tijdschrift voor Onderwijsresearch*, 18(1), 306-316.
- Helder, A., Van Leijenhof, L., & van den Broek, P. (2016). Coherence monitoring by good and poor comprehenders in elementary school: Comparing offline and online measures. *Learning and Individual Differences*, 48(1), 17-23.
<https://doi.org/10.1016/j.lindif.2016.02.008>
- Hiebert, E.H., & Mesmer, H.A.E. (2013). Upping the ante of text complexity in the Common Core State Standards: Examining its potential impact on young readers. *Educational Researcher*, 42(1), 44-51.
<https://doi.org/10.3102/0013189X12459802>
- Hoch, M.L., & McNally, J. (2020). *Empowering readers: Integrated strategies to comprehend expository texts*. Rowman and Littlefield.
- *Hoffman, K.F. (2010). *The impact of graphic organizer and metacognitive monitoring instruction on expository science text comprehension in fifth grade students* [Doctoral dissertation, North Carolina State University].
<https://repository.lib.ncsu.edu/handle/1840.16/6198>
- Hogan, T., Bridges, M.S., Justice, L.M., & Cain, K. (2011). Increasing higher-level language skills to improve reading comprehension. *Focus on Exceptional Children*, 44(3), 1-20. <https://doi.org/10.17161/foec.v44i3.6688>
- Hoogeveen, M. (2018). *Het leesonderwijs in primair en voortgezet onderwijs: Een stand van zaken en curriculaire aanbevelingen* [Reading instruction in primary and secondary education: State of affairs and curricular recommendations]. SLO.
- Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127-160. <https://doi.org/10.1007/BF00401799>
- Houtveen, A.A.M., & van de Grift, W. (2007). Effects of metacognitive strategy instruction and instruction time on reading comprehension. *School Effectiveness and School Improvement*, 18(2), 173-190.
<https://doi.org/10.1080/09243450601058717>
- Houtveen, A.A.M., van Steensel, R., & de la Rie, S. (2019). *De vele kanten van leesbegrip: Literatuurstudie naar onderwijs in begrijpend lezen in opdracht van het Nationaal Regieorgaan Onderwijsonderzoek en de Inspectie van het Onderwijs* [The many facets of reading comprehension: Literature review on reading comprehension instruction commissioned by the NRO and the Dutch Inspectorate of Education]. Erasmus University Rotterdam.
<https://www.nro.nl/sites/nro/files/migrate/de-vele-kanten-van-leesbegrip.pdf>
- Hox, J.J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

- Hudson, A.K. (2021). *Teachers' knowledge of reading comprehension, classroom practice, and student reading comprehension growth* (Publication No. 29241860) [Doctoral dissertation, Texas A&M University]. ProQuest Dissertations.
- Hudson, A.K., Moore, K.A., Han, B., Wee Koh, P., Binks-Cantrell, E., & Malatesha Joshi, R. (2021). Elementary teachers' knowledge of foundational literacy skills: A critical piece of the puzzle in the science of reading. *Reading Research Quarterly*, 56(1), 287-315. <https://doi.org/10.1002/rrq.408>
- Hughes, S., & Lewis, H. (2020). Tensions in current curriculum reform and the development of teachers' professional autonomy. *The Curriculum Journal*, 31(2), 290-302. <https://doi.org/10.1002/curj.25>
- Hultén, M., & Björkholm, E. (2016). Epistemic habits: primary school teachers' development of pedagogical content knowledge (PCK) in a design-based research project. *International Journal of Technology and Design Education*, 26(1), 335-351. <https://doi.org/10.1007/s10798-015-9320-5>
- Hwang, H., Cabell, S.Q., & Joyner, R.E. (2020). Effects of integrated literacy and content-area instruction on vocabulary and comprehension in the elementary years: A meta-analysis. *Scientific Studies of Reading*, 26(3), 223-249. <https://doi.org/10.1080/10888438.2021.1954005>
- Idol, L., & Croll, V.J. (1987). Story-mapping training as a means of improving reading comprehension. *Learning Disability Quarterly*, 10(3), 214-229. <https://doi.org/10.2307/1510494>
- Ishikawa, K. (1971). *Guide to quality control*. Asian Productivity Association.
- Janssen, F., de Graaff, R., Savelsbergh, E., & Wilschut, A. (2020). De leraar als vakdidactisch expert [The teacher as a subject-specific pedagogical expert]. In J. Dengerink, J. van der Meij, & J. Onstenk (Eds.), *Leraar: een professie met perspectief 1: Verbreding en verdieping in het beroep* (pp. 37-53). Ten Brink.
- Janzen, J., & Stoller, F.L. (1998). Integrating strategic reading into L2 instruction. *Reading in a Foreign Language*, 12(1), 251-269. <https://doi.org/10.10125/66962>
- Jiang, X., & Grabe, W. (2007). Graphic organizers in reading instruction: Research findings and issues. *Reading in a Foreign Language*, 19(1), 34-55. <https://doi.org/10.3968/9089>
- Jitendra, K., Chard, D., Hoppes, M.K., Renouf, K.M., & Gardill, C.A. (2001). An evaluation of main idea strategy instruction in four commercial reading programs: Implications for students with learning problems. *Reading and Writing Quarterly*, 17(1), 53-73. <https://doi.org/10.1080/105735601455738>
- Johnson, D.W., & Johnson, F.P. (2017). *Joining together: Group theory and group skills* (12th Ed.). Pearson.
- Johnson, D.W., & Johnson, R.T. (1984). *Circles of learning*. Association for Supervision and Curriculum Development.
- Jones, B.F., Pierce, J., & Hunter, B. (1988). Teaching students to construct graphic representations. *Educational Leadership*, 46(4), 20-25.

- Kafyulilo, A., Fisser, P., & Voogt, J. (2016). Teacher design in teams as a professional development arrangement for developing technology integration knowledge and skills of science teachers in Tanzania. *Education and Information Technologies*, 21(1), 301-318. <https://doi.org/10.1007/s10639-014-9321-0>
- Keehnen, T., Braaksma, M., & Boer, M.D. (2015). Leren door zien lezen. Observerend leren bij leesvaardigheid in 3 vwo [Learning by observing readers: Observational learning for reading proficiency in ninth grade]. *Levende Talen Tijdschrift*, 16(1), 34-41. <https://lt-tijdschriften.nl/ojs/index.php/ltt/article/view/947>
- Keenan, J.M., Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300. <https://doi.org/10.1080/10888430802132279>
- Kelly, C.H. (2019). A backwards glance: How 20th century reading research and practices have shaped current instructional approaches. In C.K. Irish (Ed.), *Making literacy connections* (pp. 4-12), University of Mary Washington.
- Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, 35(7), 1567-1577. <https://doi.org/10.3758/BF03193491>
- Kim, S.I., & van Dusen, L.M. (1998). The role of prior knowledge and elaboration in text comprehension and memory: A comparison of self-generated elaboration and text-provided elaboration and text provided elaboration. *American Journal of Psychology*, 111(3), 353-378. <https://doi.org/10.2307/1423446>
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A Construction-Integration model. *Psychological Review*, 95(2), 163-182. <https://doi.org/10.1037/0033-295X.95.2.163>
- Kintsch, W. (2004). The Construction-Integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270-1328). International Reading Association. <https://doi.org/10.1598/0872075028.46>
- Kintsch, W. (2013). Revisiting the construction-integration model of text comprehension and its implications for instruction. In D.E. Alvermann, N.J. Unrau, & R.B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 807-839). International Reading Association.
- Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394. <https://doi.org/10.1037/0033-295X.85.5.363>
- Kirkpatrick, L.C., & Klein, P.D. (2009). Planning text structure as a way to improve students' writing from sources in the compare-contrast genre. *Learning and Instruction*, 19(4), 309-321. <https://doi.org/10.1016/j.learninstruc.2008.06.001>

- Koenig, K.Q. (2018). *How do upper elementary teachers teach informational text comprehension strategies?* [Doctoral dissertation, University of Nevada]. <https://scholarworks.unr.edu/handle/11714/3419>
- Kooiker-den Boer, H.S., Sanders, T.J.M., & Evers-Vermeul, J. (2019). Tekststructuur in de Kennisbasis Nederlandse taal [Text structure in the Knowledge base for Dutch language]. *Orthopedagogiek: Onderzoek en Praktijk*, 58(11-12), 291-309. <https://orthopedagogiek.eu/article/tekststructuur-in-de-kennisbasis-nederlandse-taal/>
- Kooiker-den Boer, H.S., Sanders, T.J.M., & Evers-Vermeul, J. (to appear). Teaching text structure in science education: What opportunities do textbooks offer? *Dutch Journal of Applied Linguistics*.
- Koster, M.P., Tribushinina, E., de Jong, P., & van den Bergh, H.H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research*, 7(2), 299-324. <https://doi.org/10.17239/jowr-2015.07.02.2>
- Kostons, D., Donker, A.S., & Opdenakker, M.C. (2014). *Zelfgestuurd leren in de onderwijspraktijk: Een kennisbasis voor effectieve strategie-instructie* [Self-regulated learning in educational practice: A knowledge base for effective strategy instruction]. GION.
- Kraal, A., Koorneef, A.W., Saab, N., & van den Broek, P.W. (2018). Processing of expository and narrative texts by low-and high-comprehending children. *Reading and Writing*, 31(9), 2017-2040. <https://doi.org/10.1007/s11145-017-9789-2>
- Kucan, L., Hapgood, S., & Sullivan Palincsar, A. (2011). Teachers' specialized knowledge for supporting student comprehension in text-based discussions. *The Elementary School Journal*, 112(1), 61-82. <https://doi.org/10.1086/660689>
- Kuhlemeier, H., Jolink, A., Jongen, I., Krämer, I., Hemker, B., & Becher, T. (2014). *Balans van de leesvaardigheid in het basis- en speciaal basisonderwijs 2: Uitkomsten van de peiling in het schooljaar 2011-2012 in groep 8, groep 5 en de eindgroep van het SBO. PPON serie 54*. [Inventory of reading proficiency in regular primary education and primary education for special needs 2: Outcomes of the 2011-2012 test in grades 3 and 6. PPON series 54]. Cito.
- Kuhn, M.R. (2020). Whole class or small group fluency instruction: A tutorial of four effective approaches. *Education in Science*, 10(145), 1-11. <https://www.mdpi.com/2227-7102/10/5/145>
- Kuorikoski, J., & Marchionni, C. (2016). Evidential diversity and the triangulation of phenomena. *Philosophy of Science*, 83(2), 227-247. <https://doi.org/10.1086/684960>
- Kwint, P., & Peters, R. (2021, April 13). Motie over alleen inzetten op wetenschappelijk bewezen leesmethodes [Motion on investing in scientifically proven teaching materials for reading]. https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2021Z05863&did=2021D12929

- Land, J.F.H. (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van vmboleerlingen* [Less-skilled readers, well-built texts? Effects of text and reader characteristics on text comprehension and text appreciation]. Eburon.
- Leslie, L., & Caldwell, J.S. (2009). Formal and informal comprehension assessment. In S. Israel & G. Duffy (Eds.), *Handbook of reading comprehension* (pp. 403-427). Erlbaum.
- Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior*, 25(2), 284-289.
<https://doi.org/10.1016/j.chb.2008.12.010>
- Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic Inquiry*. Sage Publications.
- Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S., & Busick, M.D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms (NCSE 2013-3000)*. U.S. Government Printing Office.
- Maerten-Rivera, J., Ahn, S., Lanier, K., Diaz, J., & Lee, O. (2016). Effect of a multiyear intervention on science achievement of all students including English language learners. *The Elementary School Journal*, 116(4), 600-624.
<https://doi.org/10.1086/686250>
- Malone, L.D., & Mastropieri, M. (1992). Reading comprehension instruction: Summarization and self-monitoring training for students with learning disabilities. *Exceptional Children*, 58(3), 270-279.
<https://doi.org/10.1177/0014402991058003>
- Mandler, J.M., & Johnson, N.S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9(1), 111-151.
[https://doi.org/10.1016/0010-0285\(77\)90006-8](https://doi.org/10.1016/0010-0285(77)90006-8)
- Manoli, P., & Papadopoulou, M. (2012). Graphic organizers as a reading strategy: Research findings and issues. *Creative Education*, 3(3), 348-356.
<https://doi.org/10.4236/ce.2012.33055>
- Marsh, R.J., Cumming, T.M., Randolph, J.J., & Michaels, S. (2021). Updated meta-analysis of the research on response cards. *Journal of Behavioral Education*. Advance online publication. <https://doi.org/10.1007/s10864-021-09463-0>
- Mayer, R.E. (2003). *Learning and instruction*. Prentice-Hall.
- *McDermott, M.B. (1991). *Teaching top level structure as an aid for reading comprehension of expository prose for fourth-grade pupils* [Unpublished doctoral dissertation]. Hofstra University.
- McKenney, S.E., Kali, Y., Markauskaite, L., & Voogt, J. (2015). Teacher design knowledge for technology enhanced learning: an ecological framework for

- investigating assets and needs. *Instructional Science*, 43(2), 181-202.
<https://doi.org/10.1007/s11251-014-9337-2>
- McKenney, S.E., & Reeves, T.C. (2018). *Conducting educational design research*. Taylor & Francis.
- McKeown, M.G., Beck, I.L., & Blake, R.G.K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44(3), 218-253. <https://doi.org/10.1598/RRQ.44.3.1>
- *McLaughlin, E.M. (1990). *Effects of graphic organizers and levels of text difficulty on less-proficient fifth-grade readers' comprehension of expository text* [Unpublished doctoral dissertation]. University of Maryland].
- McNeil, J., & Donant, L. (1982). Summarization strategy for improving reading comprehension. In J.A. Niles, & L.A. Harris (Eds.), *New inquiries in reading research and instruction* (pp. 215-219). National Reading Conference.
- McVee, M.B., Shanahan, L.E., Hayden, H.E., Boyd, F.B., Pearson, P.D., & Reichenberg, J. (2018). *Critical reflective inquiry using the Gradual Release of Responsibility Model*. Routledge.
- Meelissen, M.R.M., Netten, A., Drent, M., Punter, R.A., Droop, M., & Verhoeven, L. (2012). *PIRLS- en TIMSS-2011. Trends in leerprestaties in lezen, rekenen en natuuronderwijs* [PIRLS and TIMSS 2011. Performance trends in reading, arithmetic, and science]. Radboud Universiteit and Universiteit Twente.
- Meyer, B.J.F. (1975). Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Reading Behavior*, 7(1), 7-47.
<https://doi.org/10.1080/10862967509547120>
- Meyer, B.J.F., Brandt, D.M., & Bluth, G.J. (1980). Use of top-level structure in text: Key for reading comprehension of 9th grade students. *Reading Research Quarterly*, 16(1), 72-103. <https://doi.org/10.2307/747349>
- Meyer, B.J.F., & Freedle, R.O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, 21(1), 121-143.
<https://doi.org/10.3102/00028312021001121>
- *Meyer, B.J.F., Middlemiss, W., Theodorou, E., Brezinski, K.L., McDougall, J., & Bartlett, B.J. (2002). Effects of structure strategy instruction delivered to fifth-grade children using the Internet with and without the aid of older adult tutors. *Journal of Educational Psychology*, 94(3), 486-519.
<https://doi.org/10.1037/0022-0663.94.3.486>
- Meyer, B.J.F., & Ray, M.N. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education*, 4(1), 127-152. <https://www.iejee.com/index.php/IEJEE/article/view/217>
- Meyer, B.J.F., & Rice, G.E. (1984). The structure of text. In P.D. Pearson (Ed.), *Handbook of reading research* (pp. 319-351). New York: Longman.
- Meyer, B.J.F., Wijekumar, K.K., & Lei, P. (2018). Comparative signaling generated for expository texts by 4th to 8th graders: Variations by text structure strategy

- instruction, comprehension skill, and signal word. *Reading and Writing*, 31(9), 1-32. <https://doi.org/10.1007/s11145-018-9871-4>
- *Meyer, B.J.F., Wijekumar, K.K., & Lin, Y.C. (2011). Individualizing a web-based structure strategy intervention for fifth graders' comprehension of nonfiction. *Journal of Educational Psychology*, 103(1), 140-168. <https://doi.org/0.1037/a0021606>
- Meyer, B.J.F., Wijekumar, K.K., Middlemiss, W., Higley, K., Lei, P.W., Meier, C., & Spielvogel, J. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly*, 45(1), 62-92. <https://doi.org/10.1598/RRQ.45.1.4>
- Miles, M.B., Huberman, A.M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage Publications.
- Miller S., Biggart A., Sloan S., O'Hare L. (2017). *Success for All: Evaluation report and executive summary*. Education Endowment Foundation.
- Ministerie OCW (2006). *Kerndoelen primair onderwijs* [Core objectives primary education]. Ministry of Education, Delta Hage.
- Miyatsu, T., Nguyen, K., & McDaniel, M.A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science*, 13(3), 390-407. <https://doi.org/10.1177/1745691617710510>
- Moelands, F., Jongen, I. van der Schoot, F., & Hemker, B. (2007). *Balans over leesstrategieën in het primair onderwijs. PPON-reeks 35* [Overview of reading strategies in primary education]. Cito.
- Moore, D.W., & Readence, J.E. (1984). A quantitative and qualitative review of graphic organizer research. *The Journal of Educational Research*, 78(1), 11-17. <https://doi.org/0.1080/00220671.1984.10885564>
- *Moore, S.R. (1996, April). *Collaboration and the reading-writing relationship: Implications for building schemata for expository text*. Paper presented at the 1996 Annual Meeting of the American Educational Research Association.
- Moss, B. (2004). Teaching expository text structures through information trade book retellings. *The Reading Teacher*, 57(8), 710-718. <https://www.jstor.org/stable/20205422>
- Muijselaar, M., Swart, N., Steenbeek-Planting, E., Droop, D., Verhoeven, L., & de Jong, P. (2018). The effect of a strategy training on reading comprehension in fourth-grade students. *The Journal of Educational Research*, 111(6), 690-703. <https://doi.org/10.1080/00220671.2017.1396439>
- Mullender-Wijnsma, M., Veldman, M., de Boer, H., van Kuijk, M., & Bosker, R. (2020). *Implementatie en effecten van Success for All in Nederland* [Implementation and effects of Success for All in the Netherlands]. GION: Groningen. <https://successforall-nederland.nl/publicaties/>

- Mullis, I.V.S., Martin, M.O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. TIMSS & PIRLS International Study Center.
- Mullis, I.V.S., Martin, M.O., Goh, S., & Prendergast, C.S. (2016). *PIRLS-2016 Encyclopedia*. TIMSS and PIRLS International Study Centre.
<http://pirls2016.org/wp-content/uploads/encyclopedia-pirls/downloadcenter/>
- Nash-Ditzel, S. (2014). Metacognitive reading strategies can improve self-regulation. *Journal of College Reading and Writing, 40*(2), 45-62.
<https://doi.org/10.1080/10790195.2010.10850330>
- Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*(3), 359-370. <https://doi.org/10.1111/j.2044-8279.1997.tb01250.x>
- Nikula, T., Dalton-Puffer, C., & Llinares, A. (2013). CLIL classroom discourse: Research from Europe. *Journal of Immersion and Content-Based Language Education, 1*(1), 70-100. <https://doi.org/10.1075/jicb.1.1.04nik>
- Nolte, R.Y., & Singer, H. (1985). Active comprehension: Teaching a process of reading comprehension and its effects on reading achievement. *The Reading Teacher, 39*(1), 24-31. <https://www.jstor.org/stable/20198994>
- Nutley, S. M., Walter, I., & Davies, H. T. (2007). *Using evidence: How research can inform public services*. Policy Press.
- Oakhill, J., Cain, K., & Elbro, C. (2015). *Understanding and teaching reading comprehension: A handbook*. Routledge.
- *Ocasio, T.L. (2006). *A comparison of two instructional programs to develop strategies to improve reading comprehension* (Publication no. 3209082) [Doctoral dissertation, Widener University]. ProQuest Dissertations.
- Odom, S.L. (2009). The tie that binds: Evidence based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education, 29*(1), 53-61. <https://doi.org/10.1177/0271121408329171>
- Ogle, D., & Blachowicz, C.L. (2002). Beyond literature circles: Helping students comprehend informational texts. In C.C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 259-274). Guilford.
- Okkinga, M., van Gelderen, A.J., van Schooten, E., van Steensel, R., & Slegers, P.J.C. (2021). Implementation quality of principles of reciprocal teaching in whole-classroom settings: a two-year study with low-achieving adolescents. *Reading Psychology, 42*(4), 323-363. <https://doi.org/10.1080/02702711.2021.1887019>
- Okkinga, M., van Steensel, R., van Gelderen, A.J., & Slegers, P. J. (2018b). Effects of reciprocal teaching on reading comprehension of low-achieving adolescents. The importance of specific teacher skills. *Journal of Research in Reading, 41*(1), 20-41. <https://doi.org/10.1111/1467-9817.12082>
- Okkinga, M., van Steensel, R., van Gelderen, A.J., van Schooten, E., Slegers, P.J., & Arends, L.R. (2018a). Effectiveness of reading-strategy interventions in whole

- classrooms: A meta-analysis. *Educational Psychology Review*, 30(4), 1215-1239.
<https://doi.org/10.1007/s10648-018-9445-7>
- Oosterloo, A. (2010). De begrijpend leesmethode loslaten [Letting go of textbooks for reading comprehension]. *Tijdschrift Taal voor Opleiders en Onderwijsadviseurs*, 1, 19-25.
- Ormel, B.J., Roblin, N., McKenney, S.E., Voogt, J.M., & Pieters, J.M. (2012). Research-practice interactions as reported in recent design studies: Still promising, still hazy. *Educational Technology Research and Development*, 60(6), 967-986.
<https://doi.org/10.1007/s11423-012-9261-6>
- Palinscar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension monitoring activities. *Cognition and Instruction*, 1(2), 117-175.
https://doi.org/10.1207/s1532690xci0102_1
- Paris, S.G., Lipson, M.Y., & Wixson, K.K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8(3), 293-316.
[https://doi.org/10.1016/0361-476X\(83\)90018-8](https://doi.org/10.1016/0361-476X(83)90018-8)
- Patton, S.A., Fuchs, D., Hendricks, E.L., Pennell, A.J., Walsh, M.E., Fuchs, L.S., Tracy, W.Z., & Haga, L.Y. (2022). An experimental study to strengthen students' comprehension of informational texts: Is teaching for transfer important? *Learning Disabilities Research and Practice*, 37(2), 124-139.
<https://doi.org/10.1111/ldrp.12276>
- Pearson, P.D., & Gallagher, M.C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology*, 8(3), 317-344.
[https://doi.org/10.1016/0361-476X\(83\)90019-X](https://doi.org/10.1016/0361-476X(83)90019-X)
- Penuel, W.R., Phillips, R.S., & Harris, C.J. (2014). Analysing teachers' curriculum implementation from integrity and actor-oriented perspectives. *Journal of Curriculum Studies*, 46(6), 751-777. <https://doi.org/10.1080/00220272.2014.921841>
- Pereira, C., & Nicolaas, M. (2019). *Effectief onderwijs in begrijpend lezen: Acties voor beter leesbegrip en meer leesmotivatie* [Effective reading comprehension instruction: Actions for better comprehension and more reading motivation]. Nederlandse Taalunie. <https://taalunie.org/publicaties/17/effectief-onderwijs-in-begrijpend-lezen>
- Peti-Stantić, A., Keresteš, G., & Gnjiđić, V. (2021). Can textbook analysis help us understand why Croatian students seldom read their textbooks? *Technology, Knowledge and Learning*, 26(1), 293-310. <https://doi.org/10.1007/s10758-020-09485-z>
- Piasta, S.B., Connor, C.M., Fishman, B.J., & Morrison, F.J. (2009). Teachers' knowledge of literacy concepts, classroom practices, and student reading growth. *Scientific Studies of Reading*, 13(3), 224-248. <https://doi.org/10.1080/10888430902851364>

- Piper, S.A. (2019). *Upper elementary teachers' use of pedagogical content knowledge with nonfiction reading instruction* [Doctoral dissertation, Walden University]. <https://core.ac.uk/download/pdf/223224248.pdf>
- *Ponce, H.R., López, M.J., & Mayer, R.E. (2012). Instructional effectiveness of a computer-supported program for teaching reading comprehension strategies. *Computers & Education*, 59(4), 1170-1183. <https://doi.org/10.1016/j.compedu.2012.05.013>
- Pressley, M., El-Dinary, P.B., Gaskins, I., Schuder, T., Bergman, J.L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92(5), 513-555. <https://doi.org/10.1086/461705>
- Pulles, M., Berenst, J., de Glopper, K., & Koole, T. (2022). Children's discussions about texts: Integrating and evaluating practices. *Linguistics and Education*, 69, 1-16. <https://doi.org/10.1016/j.linged.2022.101051>
- Puzio, K., & Colby, G.T. (2013). Cooperative learning and literacy: A meta-analytic review. *Journal of Research on Educational Effectiveness*, 6(4), 339-360. <https://doi.org/10.1080/19345747.2013.775683>
- Pyle, N., Vasquez, A.C., Lignugaris/Kraft, B., Gillam, S.L., Reutzel, D.R., Olszewski, A., Segura, H., Hartzheim, D., Laing, W., & Pyle, D. (2017). Effects of expository text structure interventions on comprehension: A meta-analysis. *Reading Research Quarterly*, 52(4), 469-501. <https://doi.org/10.1002/rrq.179>
- Raad voor Cultuur & Onderwijsraad (2019). *Lees! Een oproep voor een leesoffensief*. <https://www.onderwijsraad.nl/publicaties/adviezen/2019/06/24/leesadvies>
- *Raphael, T.E., Englert, C.S., & Kirschner, B.W. (1986). *The impact of text structure instruction and social context on students' comprehension and production of expository text* (Research Series No. 177). Institute for Research on Teaching, Michigan State University.
- *Raphael, T.E., & Kirschner, B.M. (1985). *The effects of instruction in compare/contrast text structure on sixth-grade students' reading comprehension and writing products* (Research Series No. 161). Institute for Research on Teaching, Michigan State University.
- Rapp, D.N., van den Broek, P., McMaster, K.L., Kendeou, P., & Espin, C.A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11(4), 289-312. <https://doi.org/10.1080/10888430701530417>
- Read, S., Reutzel, D.R., & Fawson, P.C. (2008). Do you want to know what I learned? Using informational trade books as models to teach text structure. *Early Childhood Education Journal*, 36(3), 213-219. <https://doi.org/10.1007/s10643-008-0273-0>
- Reutzel, D.R., Jones, C.D., Clark, S.K., & Kumar, T. (2016). The Informational Text Structure Survey (ITS2): An exploration of primary grade teachers' sensitivity to

- text structure in young children's informational texts. *The Journal of Educational Research*, 109(1), 81-98. <https://doi.org/10.1080/00220671.2014.918927>
- Reutzel, D.R., Smith, J.A., & Fawson, P.C. (2005). An evaluation of two approaches for teaching reading comprehension strategies in the primary years using science information texts. *Early Childhood Research Quarterly*, 20(3), 276-305. <https://doi.org/10.1016/j.ecresq.2005.07.002>
- Rijlaarsdam, G. (2005). Observerend leren: een kernactiviteit in taalvaardigheidsonderwijs: Ontwerpadvies uit onderzoek verkregen [Observational learning: design principles derived from research]. *Levende Talen Tijdschrift*, 6(4), 10-28. <https://lt-tijdschriften.nl/ojs/index.php/lt/article/view/488>
- Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., Broekkamp, H., & van den Bergh, H. (2005). Psychology and the teaching of writing in 8000 and some words. In *BJEP Monograph Series II, Number 3-Pedagogy-Teaching for Learning* (Vol. 127:153, pp. 127-153). British Psychological Society.
- Robinson, D.H., & Molina, E. (2002). The relative involvement of visual and auditory working memory when studying adjunct displays. *Contemporary Educational Psychology*, 27(1), 118-131. <https://doi.org/10.1006/ceps.2001.1081>
- Robitaille, D.F., Schmidt, W.H., Raizen, S.A., McKnight, C.C., Britton, E.D., & Nicol, C. (1993). *Curriculum Frameworks for Mathematics and Science*. Pacific Educational Press.
- Rogiers, A., Merchie, E., & Van Keer, H. (2019). Learner profiles in secondary education: occurrence and relationship with performance and student characteristics. *Journal of Educational Research*, 112(3), 385-396. <https://doi.org/10.1080/00220671.2018.1538093>
- Rogiers, A., Merchie, E., & Van Keer, H. (2021). Illuminating learning from informative texts in secondary education: A switching replication design study. *Contemporary Educational Psychology*, 64, 101946. <https://doi.org/10.1016/j.cedpsych.2021.101946>
- Rooijackers, P., van Silfhout, G., Schuurs, U., Mulders, I., & van den Bergh, H. (2020). Lezen en antwoorden bij teksten met vragen. Een cross-sectionele eye-trackstudie onder vwo-leerlingen [Reading and answering texts with questions: a cross-sectional eye-tracking study among pre-university students]. *Pedagogische Studiën*, 93(4), 42-58.
- Rooijackers, P., van Silfhout, G., Schuurs, U., Mulders, I., & van den Bergh, H. (2021). De relatie tussen het vooraf lezen van teksten en het beantwoorden van begripsvragen: twee experimenten onder 5-vwo-leerlingen. *Pedagogische Studiën*, 98(5), 320-349.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479-530. <https://doi.org/10.3102/00346543064004479>

- Rouffet, C., van Beuningen, C., & de Graaff, R. (2022). Constructive alignment in foreign language curricula: an exploration of teaching and assessment practices in Dutch secondary education. *The Language Learning Journal*, 1-15. <https://doi.org/10.1080/09571736.2022.2025542>
- Sáenz, L.M., & Fuchs, L.S. (2002). Examining the reading difficulty of secondary students with learning disabilities: Expository versus narrative text. *Remedial and Special Education*, 23(1), 31-41. <https://doi.org/10.1177/074193250202300105>
- Sánchez, E., & García, J.R. (2009). The relation of knowledge of textual integration devices to expository text comprehension under different assessment conditions. *Reading and Writing*, 22(9), 1081-1108. <https://doi.org/10.1007/s11145-008-9145-7>
- Sanders, T.J.M., & Noordman, L.G.M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), 37-60. https://doi.org/10.1207/S15326950dp2901_3
- Sanders, T.J.M., & Spooren, W.P.M.S. (2007). Discourse and text structure. In D. Geeraerts & H. Cuykens (Eds.), *Handbook of cognitive linguistics* (pp. 916-941). Oxford University Press.
- Sangers, N.L., Evers-Vermeul, J., Hoeken, H., & Sanders, T.J.M. (2021). Narrative elements in expository texts: a corpus study of educational textbooks. *Dialogue and Discourse*, 12(2), 115-144. <https://doi.org/10.5210/dad.2021.204>
- Scarborough, H. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S.B. Neuman & D.K. Dickinson (Eds.), *Handbook of early literacy* (pp. 97-110). Guilford.
- Scheerens, J. (2017). Recapitalization, implications for educational policy, practice and future research. In J. Scheerens (Ed.), *Opportunity to learn, curriculum alignment, and test preparation: A research review* (pp. 121-135). Oberon.
- Scheerens, J., & Blömeke, S. (2016). Integrating teacher education effectiveness research into educational effectiveness models. *Educational Research Review*, 18(1), 70-87. <https://doi.org/10.1016/j.edurev.2016.03.002>
- Scheltinga, F., van der Maas, M., Bontje, J., & van der Hoeven, J. (2013). *Leerlijn, leesmethoden en onderwijsaanbod: Begrijpend lezen onderzocht* [Curricular structure, teaching materials, and instruction: Investigating reading comprehension]. KPC Groep.
- Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistics perspective*. Lawrence Erlbaum.
- Schnabel, P., Dam, G., Douma, T., van Eijk, R., Tabarki, F., van der Touw, A., Verweij, J., & Visser, M. (2016). *Ons onderwijs: 2032. Eindadvies* [Our education: 2032. Final advice]. Platform Onderwijs 2032.
- Schneider, R.M. (2013). Opportunities for teacher learning during enactment of inquiry science curriculum materials: Exploring the potential for teacher educative materials. *Journal of Science Teacher Education*, 24(2), 323-346. <https://doi.org/10.1007/s10972-012-9309-9>

- Schneider, R.M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of teacher enactments in reform classrooms. *Journal of Research in Science Teaching*, 42(3), 283-312. <https://doi.org/10.1002/tea.20055>
- Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8, and 10. *Language Learning*, 48(1), 71-106. <https://doi.org/10.1111/1467-9922.00033>
- Schutz, K.M., & Rainey, E.C. (2020). Making sense of modeling in elementary literacy instruction. *The Reading Teacher*, 73(4), 443-451. <https://doi.org/10.1002/trtr.1863>
- *Scott, D.B. (2011). *Explicit instruction on rhetorical patterns and student-constructed graphic organizers: The impact on sixth-grade students' comprehension of social studies text* [Doctoral dissertation, University of Maryland]. <https://drum.lib.umd.edu/handle/1903/11736>
- Seifert, S. (2021). Is reading comprehension taken for granted? An analysis of Austrian textbooks in fourth and sixth grade. *Technology, Knowledge and Learning*, 26(1), 383-405. <https://doi.org/10.1007/s10758-021-09490-w>
- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Shanahan, L.E., Tochelli-Ward, A.L., & Rinker, T.W. (2019). The ebb and flow of scaffolding: Thinking flexibly about the gradual release of responsibility during explicit strategy instruction. In M.B. McVee, E. Ortlieb, J.S. Reichenberg, & P.D. Pearson (Eds.), *The Gradual Release of Responsibility in literacy research and practice* (Vol. 10, pp. 53-66). Emerald Publishing.
- Shanahan, T., Callison, K., Carriere, C., Duke, N.K., Pearson, P.D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: IES practice guide. NCEE 2010-4038*. What Works Clearinghouse.
- *Short, E.J., & Ryan, E.B. (1984). Metacognitive differences between skilled and less skilled readers: Remediating deficits through story grammar and attribution training. *Journal of Educational Psychology*, 76(2), 225-235. <https://doi.org/10.1037/0022-0663.76.2.225>
- Simpson, M.L., & Nist, S.L. (2000). An update on strategic learning: It's more than textbook reading strategies. *Journal of Adolescent and Adult Literacy*, 43(6), 528-541. <https://www.jstor.org/stable/40016831>
- Sinatra, G.M., Brown, K.J., & Reynolds, R.E. (2002). Implications of cognitive resource allocation for comprehension strategies instruction. In C.C. Block & M. Presley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 62-76). Guilford.
- Singer, H., & Donlan, D. (1980). *Reading and learning from text*. Little Brown.

- Slavin, R.E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, 57(3), 293-336. <https://doi.org/10.3102/00346543057003293>
- Slavin, R.E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, 79(4), 1391-1466. <https://www.jstor.org/stable/40469101>
- Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, 42(3), 214-240. <https://doi.org/10.1080/02702711.2021.1888348>
- Smith, S.J., & Okolo, C. (2010). Response to Intervention and evidence-based practices: Where does technology fit? *Learning Disability Quarterly*, 33(4), 257-272. <https://doi.org/10.1177/073194871003300404>
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Corporation. <http://www.rand.org/content/dam/rand/pubs/monographreports/2005/MR1465.pdf>
- Soodla, P., Jögi, A., & Kikas, E. (2017). Relationships between teachers' metacognitive knowledge and students' metacognitive knowledge and reading achievement. *European Journal of Psychology of Education*, 32(2), 201-218. <https://www.jstor.org/stable/44951885>
- Soto, C., Gutiérrez de Blume, A.P., Jacovina, M., McNamara, D., Benson, N., & Riffo, B. (2019). Reading comprehension and metacognition: The importance of inferential skills. *Cogent Education*, 6(1), 1-20. <https://doi.org/10.1080/2331186X.2019.1565067>
- *Spires, H.A., Gallini, J., & Riggsbee, J. (1992). Effects of schema-based and text structure-based cues on expository prose comprehension in fourth graders. *Journal of Experimental Education*, 60(4), 307-320. <https://doi.org/10.1080/00220973.1992.9943868>
- Spiegel, G.F., & Barufaldi, J.P. (1994). The effects of a combination of text structure awareness and graphic post-organizers on recall and retention of science knowledge. *Journal of Research in Science Teaching*, 31(9), 913-932. <https://doi.org/10.1002/tea.3660310907>
- Spörer, N., Brunstein, J. C., & Kieschke, U.L.F. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and instruction*, 19(3), 272-286. <https://doi.org/10.1016/j.learninstruc.2008.05.003>
- Stahl, S.A., & Fairbanks, M.M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, 56(1), 72-110. <https://doi.org/10.3102/00346543056001072>

- Stanovich, K.E., & Stanovich, P.J. (1995). How research might inform the debate about early reading acquisition. *Journal of Research in Reading, 18*(2), 87-105. <https://doi.org/10.1111/j.1467-9817.1995.tb00075.x>
- *Stevens, E.A. (2018). *The effects of a text structure and paraphrasing intervention on the main idea generation and reading comprehension of struggling readers in grades 4 and 5* [Doctoral dissertation, University of Texas]. <https://repositories.lib.utexas.edu/handle/2152/69020>
- Stevens, E.A., & Vaughn, S. (2021). Using paraphrasing and text structure instruction to support main idea generation. *Teaching Exceptional Children, 53*(4), 300-308. <https://doi.org/10.1177/0040059920958738>
- Stevens, R.J., Slavin, R.E., & Farnish, A.M. (1991). The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification. *Journal of Educational Psychology, 83*(1), 8-16. <https://doi.org/10.1037/0022-0663.83.1.8>
- Stine-Morrow, E.A., Gagne, D.D., Morrow, D.G., & DeWall, B.H. (2004). Age differences in rereading. *Memory & Cognition, 32*(5), 696-710. <https://doi.org/10.3758/BF03195860>
- Strong, J.Z. (2020). Investigating a text structure intervention for reading and writing in grades 4 and 5. *Reading Research Quarterly, 55*(4), 545-551. <https://doi.org/10.1002/rrq.356>
- Stull, A.T., & Mayer, R.E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology, 99*(4), 808-820. <https://doi.org/10.1037/0022-0663.99.4.808>
- *Tackett, S.A., Patberg, J.P., & Dewitz, P. (1984). The effects of story structure instruction on the recall of poor sixth grade readers. *Reading Psychology, 5*(3), 193-201. <https://doi.org/10.1080/0270271840050303>
- *Taylor, B.M. (1982). Text structure and children's comprehension and memory for expository material. *Journal of Educational Psychology, 74*(3), 323-340. <https://doi.org/10.1037/0022-0663.74.3.323>
- *Taylor, B.M. (1985). Improving middle-grade students' reading and writing of expository text. *The Journal of Educational Research, 79*(2), 119-125. <https://doi.org/10.1080/00220671.1985.10885661>
- Taylor, B.M., & Beach, R.W. (1984). The effects of text structure instruction on middle-grade students' comprehension and production of expository text. *Reading Research Quarterly, 19*(2), 134-146. <https://doi.org/10.2307/747358>
- Taylor, B.M., Pearson, P.D., Peterson, D.S., & Rodriguez, M.C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal, 104*(1), 3-28. <https://doi.org/10.1086/499740>

- Timberlake, M.T., Thomas, A.B., & Barrett, B. (2017). The allure of simplicity: Scripted curricula and equity. *Teaching and Teacher Education*, 67(1), 46-52.
<https://doi.org/10.1016/j.tate.2017.05.017>
- *Troyer, S.J. (1994, April). *The effects of three instructional conditions in text structure on upper elementary students' reading comprehension and writing performance*. Paper presented at the 1994 Annual Meeting of the American Educational Research Association.
- Tuithof, H., van Drie, J., Bronkhorst, L., Dorsman, L., & van Tartwijk, J. (2021). Teachers' pedagogical content knowledge of two specific historical contexts captured and compared. *Educational Studies*, 1-26.
<https://doi.org/10.1080/03055698.2021.1877621>
- Turcotte, C., Giguère, M.H., & Godbout, M.J. (2015). Une approche d'enseignement des stratégies de compréhension de lecture de textes courants auprès de jeunes lecteurs à risque d'échouer [An instructional approach to teach reading strategies to young readers at risk for reading failure]. *Language & Literacy*, 17(1), 106-125. <https://doi.org/10.20360/G2SW2B>
- Turner, F.D. (2010). Evaluating the effectiveness of fluency-oriented reading instruction with increasing Black and Latino reading fluency, as compared to Asian and White second-grade students' reading fluency. *The Journal of Negro Education*, 79(2), 112-124. <https://www.jstor.org/stable/20798330>
- *Ulper, H., & Akkok, E.A. (2010). The effect of using expository text structures as a strategy on summarization skills. In L.E. Kattington (Ed.), *Handbook of curriculum development* (pp. 303-328). Nova Science.
- Valencia, S.W., Place, N.A., Martin, S.D., & Grossman, P.L. (2006). Curriculum materials for elementary reading: Shackles and scaffolds for four beginning teachers. *The Elementary School Journal*, 107(1), 93-120.
<https://doi.org/10.1086/509528>
- Valverde, G.A., Bianchi, L.J., Wolfe, R.G., Schmidt, W.H., & Houang, R.T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Kluwer Academic Publishers.
- Van den Branden, K., Van Keer, H., & Ghesquière, P. (2019). *Sleutels voor effectief begrijpend lezen. Inspiratie voor een eigentijds didactiek in het basisonderwijs* [Keys for effective reading comprehension: Inspiration for a contemporary pedagogy in primary education]. Vlaamse Onderwijsraad.
- van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, 328(5977), 453-456.
<https://www.science.org/doi/10.1126/science.1182594>
- van den Broek, P., Helder, A., Espin, C., & van der Liende, M. (2021). *Sturen op begrip: Effectief leesonderwijs in Nederland* [Fostering text comprehension: Effective reading comprehension instruction in the Netherlands]. Leiden University.
<https://www.tweedekamer.nl/sites/default/files/atoms/files/>

- 20220203_onderzoeksrapport_sturen_op_begrip_effectief_leesonderwijs_in_nederland_in_opdracht_van_commissie_voor_ocw.pdf
- van den Broek, P., & Kremer, K.E. (2000). The mind in action: What it means to comprehend during reading. In B.M. Taylor, M.F. Graves, & P. van den Broek (Eds.), *Reading for meaning: Fostering comprehension in the middle grades* (pp. 1-31). Teachers College Press; International Reading Association.
- van den Broek, P., Rapp, D.N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, 39(2-3), 299-316. https://doi.org/10.1207/s15326950dp3902&3_11
- van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The Landscape Model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98). Erlbaum.
- van der Hoeven, J., & Bontje, J. (2013). *Basisvaardigheden taal en rekenen in het hart van het onderwijs: vier jaar praktijkgericht onderzoek* [Basic skills language and arithmetic at the heart of instruction: four years of practice-oriented research]. KPC Groep.
- Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, 36(2), 299-316. <https://www.jstor.org/stable/27823607>
- van der Schoot, M., Horsley, T.M., & van Lieshout, E.C.D.M. (2010). The effects of instruction on situation model construction: An eye fixation study on text comprehension in primary school children. *Educational Psychology*, 30(7), 817-835. <https://doi.org/10.1080/01443410.2010.510600>
- van de Wetering, T., & Groenendijk, R. (2015). Van toetsgerichte lezers naar zelfstandige, analytische lezers: Geen trucje maar een vaardigheid. *Levende Talen Magazine*, 102(2), 4-8. <https://lt-tijdschriften.nl/ojs/index/ltm/article/view/929>
- van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- van Gelderen, A., & van Schooten, E.J. (2011). *Taalonderwijs: Een kwestie van ontkavelen* [Language education: A matter of decompartmentalizing]. Rotterdam University Press.
- van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22(2), 155-174. <https://doi.org/10.1007/s10648-010-9134-7>
- van Kuijk, M., Mullender-Wijnsma, M., & Bosker, R. (2021). A systematic review of studies assessing the implementation of the evidence-based whole-school

- reform 'Success for All'. *Review of Education*, 4(1), 128-163.
<https://doi.org/10.1177/20965311209615>
- van Moort, M.L., Jolles, D.D., Koornneef, A., & van den Broek, P. (2020). What you read versus what you know: Neural correlates of accessing context information and background knowledge in constructing a mental representation during reading. *Journal of Experimental Psychology: General*, 149(11), 2084-2101.
<https://doi.org/10.1037/xge0000764>
- van Silfhout, G., Evers-Vermeul, J., & Sanders, T.J.M. (2015). Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, 52(1), 47-76.
<https://doi.org/10.1080/0163853X.2014.905237>
- *Van Steenbrugge, H. (2006). *Begrijpend lezen en schematiseren: impact van schema's maken op metacognitie* [Reading comprehension and schematizing: Impact of schematizing on metacognition] (Master's thesis). Retrieved from <https://lib.ugent.be/en/catalog/rug01:001043530>
- van Weerden, J. & Hiddink, L. (2013). *Balans van het basisonderwijs. PPON: 25 jaar kwaliteit in beeld* [Balance of primary education. PPON: 25 years of educational quality in the picture]. Cito.
- Vaughn, S., Klinger, J., & Bryant, D. (2001). Collaborative strategic reading as a means to enhance peer-mediated instruction for reading comprehension and content-area learning. *Remedial and Special Education*, 22(2), 66-74.
<https://doi.org/10.1177/074193250102200201>
- Vázquez, V.P. (2014). Enhancing the quality of CLIL: Making the best of the collaboration between language teachers and content teachers. *Encuentro*, 23, 115-127.
- Veldman, M.A., Van Kuijk, M.F., Doolaard, S., & Bosker, R.J. (2020). The proof of the pudding is in the eating? Implementation of cooperative learning: Differences in teachers' attitudes and beliefs. *Teachers and Teaching*, 26(1), 103-117.
<https://doi.org/10.1080/13540602.2020.1740197>
- Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology*, 22(3), 293-301.
<https://doi.org/10.1002/acp.1417>
- *Vidal-Abarca, E. (1990). Un programa para la enseñanza de la comprensión de ideas principales de textos expositivos [A program for teaching main idea comprehension in expository texts]. *Journal for the Study of Education and Development*, 13(49), 53-71. <https://doi.org/10.1080/02103702.1990.10822257>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.
<https://doi.org/10.18637/jss.v036.i03>

- Voogt, J., Westbroek, H., Handelzalts, A., Walraven, A., McKenney, S., Pieters, J., & de Vries, B. (2011). Teacher learning in collaborative curriculum design. *Teaching and Teacher Education*, 27(8), 1235-1244. <https://doi.org/10.1016/j.tate.2011.07.003>
- *Walker, M.L. (1995). Help for the "fourth-grade slump": SRQ2R plus instruction in text structure or main idea. *Reading Horizons*, 36(1), 38-58.
https://scholarworks.wmich.edu/reading_horizons/vol36/iss1/3
- Weaver, P.A., & Dickinson, D.K. (1982). Scratching below the surface structure: Exploring the usefulness of story grammar. *Discourse Processes*, 5(3), 225-243.
<https://doi.org/10.1080/01638538209544544>
- Webb, S., Massey, D., Goggans, M., & Flajole, K. (2019). Thirty-five years of the gradual release of responsibility: Scaffolding toward complex and responsive teaching. *The Reading Teacher*, 73(1), 75-83. <https://doi.org/10.1002/trtr.1799>
- Westby, C., Culatta, B., Lawrence, B., & Hall-Kenyon, K. (2010). Summarizing expository texts. *Topics in Language Disorders*, 30(4), 275-287.
<https://doi.org/10.1097/TLD.0b013e3181ff5a88>
- Wieberdink, H., & Kuster, H. (2015). Naar een cognitief-linguïstische benadering van begrijpend lezen op de basisschool [Towards a cognitive-linguistic approach to reading comprehension in elementary school]. *Orthopedagogiek: Onderzoek en Praktijk*, 54(2), 91-104.
- Wijekumar, K.K., Beerwinkle, A.L., Harris, K.R., & Graham, S. (2019a). Etiology of teacher knowledge and instructional skills for literacy at the upper elementary grades. *Annals of Dyslexia*, 69(1), 5-20. <https://doi.org/10.1007/s11881-018-00170-6>
- Wijekumar, K.K., Beerwinkle, A.L., McKeown, D., Zhang, S., & Joshi, R.M. (2020). The GIST of the reading comprehension problem in grades 4 and 5. *Dyslexia*, 26(3), 323-340. <https://doi.org/10.1002/dys.1647>
- *Wijekumar, K.K., Meyer, B.J.F., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, 60(6), 987-1013. <https://doi.org/10.1007/s11423-012-9263-4>
- *Wijekumar, K.K., Meyer, B.J.F., & Lei, P. (2013). High-fidelity implementation of web-based intelligent tutoring system improves fourth and fifth graders content-area reading comprehension. *Computers & Education*, 68(1), 366-379.
<https://doi.org/10.1016/j.compedu.2013.05.021>
- Wijekumar, K.K., Meyer, B.J.F., Lei, P., Beerwinkle, A.L., & Joshi, M. (2019b). Supplementing teacher knowledge using web-based Intelligent Tutoring System for the Text Structure Strategy to improve content-area comprehension with fourth- and fifth-grade struggling readers. *Dyslexia*, 26(2), 120-136.
<https://doi.org/10.1002/dys.1634>

- *Wijekumar, K.K., Meyer, B.J.F., Lei, P.W., Lin, Y.C., Johnson, L.A., Spielvogel, J.A., Shurmatz, K.M., Ray, M., & Cook, M. (2014). Multisite randomized controlled trial examining intelligent tutoring of structure strategy for fifth-grade readers. *Journal of Research on Educational Effectiveness*, 7(4), 331-357. <https://doi.org/10.1080/19345747.2013.853333>
- Wijekumar, K.K., Zhang, S., Joshi, R.M., & Peti-Stantić, A. (2021). Introduction to the special issue on textbook content and organization: why it matters to reading comprehension in elementary grades? *Technology, Knowledge and Learning*, 26(1), 243-249. <https://doi.org/10.1007/s10758-021-09505-6>
- Wikeley, F., Stoll, L., Murillo, J., & de Jong, R. (2005). Evaluating effective school improvement: Case studies of programmes in eight European countries and their contribution to the effective school improvement model. *School Effectiveness and School Improvement*, 16(4), 387-405. <https://doi.org/10.1080/09243450500234617>
- Williams, J.P. (1998). Improving the comprehension of disabled readers. *Annals of Dyslexia*, 48, 213-238. <https://link.springer.com/article/10.1007/s11881-998-0010-9>
- Williams, J.P. (2005). Instruction in reading comprehension for primary-grade students: A focus on text structure. *The Journal of Special Education*, 39(1), 6-18. <https://doi.org/10.1177/002246690503900102>
- Williams, J.P. (2018). Text structure instruction: The research is moving forward. *Reading and Writing: an Interdisciplinary Journal*, 31(9), 1923-1935. <https://doi.org/10.1007/s11145-018-9909-7>
- Williams, J.P., Hall, K.M., & Lauer, K. (2004). Teaching expository text structure to young at-risk learners: Building basics of comprehension instruction. *Exceptionality*, 12(3), 129-144. https://doi.org/10.1207/s15327035ex1203_2
- Williams, J.P., Kao, J.C., Pao, L.S., Ordynans, J.G., Atkins, J.G., Cheng, R., & DeBonis, D. (2016). Close analysis of texts with structure (CATS): An intervention to teach reading comprehension to at-risk second graders. *Journal of Educational Psychology*, 108(8), 1061-1077. <https://doi.org/10.1037/edu0000117>
- Williams, J.P., Pollini, S., Nubla-Kung, A.M., Snyder, A.E., Garcia, A., Ordynans, J.G., & Atkins, J.G. (2014). An intervention to improve comprehension of cause/effect through expository text structure instruction. *Journal of Educational Psychology*, 106(1), 1-17. <https://doi.org/10.1037/a0033215>
- Wilkinson, I.A.G., & Son, E.H. (2011). A dialogic turn in research on learning and teaching to comprehend. In M.L. Kamil, P.D. Pearson, E.B. Moje & P. Afflerbach (Eds.), *Handbook of Reading Research* (pp. 359-382). Routledge.
- Winograd, P.N. (1984). Strategic difficulties in summarizing texts. *Reading Research Quarterly*, 19(4), 404-425. <https://doi.org/10.2307/747913>
- Winograd, P.N., & Hare, V.C. (1988). Direct instruction of reading comprehension strategies: The nature of teacher explanation. In C.E. Weinstein, E.T. Goetz & P.A.

- Alexander (Eds.), *Learning and Study Strategies: Issues in Assessment Instruction and Evaluation* (pp. 121-139). Academic Press.
- Wood, D., Bruner, J.S., & Ross, G. (1976). The role of tutoring in problem solving. *Child Psychology and Psychiatry*, 17(2), 89-100.
<https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Wolman, C. (1991). Sensitivity to causal cohesion in stories by children with mild mental retardation, children with learning disabilities, and children without disabilities. *The Journal of Special Education*, 25(2), 135-154.
<https://doi.org/10.1177/00224669910250020>
- Wu, Y., Barquero, L.A., Pickren, S.E., Taboada, A.M., & Cutting, L.E. (2020). The relationship between cognitive skills and reading comprehension of narrative and expository texts: A longitudinal study from Grade 1 to Grade 4. *Learning and Individual Differences*, 80, 1-14. <https://doi.org/10.1016/j.lindif.2020.101848>
- Xu, Z., Wijekumar, K.K., Wang, Q., Irey, R., & Liang, H. (2021). The effects of web-based text structure strategy instruction on adult Chinese ELLs' reading comprehension and reading strategy use. *Language Teaching Research*. Online preprint. <https://doi.org/10.1177/1362168821102230>
- Yang, Y. F. (2006). Reading strategies or comprehension monitoring strategies?. *Reading Psychology*, 27(4), 313-343. <https://doi.org/10.1080/02702710600846852>
- Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, 2(1), 33-45.
<https://doi.org/10.1002/acp.2350020105>
- Zarrati, Z., Nambiar, R.N.K., Rizan,, & Maasum, T.N.T.M. (2014). The importance of text structure awareness in promoting strategic reading among EFL readers. *Procedia: Social and Behavioral Sciences*, 118(1), 537-544.
<https://doi.org/10.1016/j.sbspro.2014.02.073>
- Zhang, S., Wijekumar, K.K., & Han, B. (2021). An analysis of Grade 4 reading textbooks used in mainland China: Do the texts and activities support higher order reading comprehension skills? *Technology, Knowledge and Learning*, 26(1), 251-291. <https://doi.org/10.1007/s10758-021-09504-7>
- Zimmerman, B., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22(1), 73-101.
<https://doi.org/10.1006/ceps.1997.0919>
- Zwaan, R.A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20(4), 920-933.
<https://doi.org/10.1037/0278-7393.20.4.920>

Appendix A: Meta-analysis (Chapter 2)

This appendix to *Chapter 2* provides an overview of our stepwise model fitting and the parameters of the final model for each of the outcome measures. Because the analysis of moderating content-related and instructional variables (models 6 and 7 for comprehension questions and recall, and models 5 and 6 for summarization) might lead to an excessive number of cells, in a second step, we compared the full-factorial model with a more economic model including only the dummy variables that seemed to matter and had sufficient independent observations in our data sets. Specifically, we first analyzed whether the immediate and delayed effects of text structure instruction were affected by structure visualizations (e.g., maps, graphic organizers – GO's) and structure-based summarization techniques (e.g., main idea identification, outlining). This did not improve the model fit: comprehension questions, $\Delta\chi^2(2) = 0.73$, $p = .69$; recall, $\Delta\chi^2(2) = 0.082$, $p = .96$; and summarization, $\Delta\chi^2(2) = 5.78$, $p = .06$.

Because the text structure interventions were very heterogeneous, we added three more specific dummy variables: a focus on paragraph structure (i.e., topic sentences, supporting details), active use of GOs, and teaching rule-based summarization. This improved the model fits for all outcome measures: comprehension questions, $\Delta\chi^2(2) = 6.57$, $p = .04$; recall, $\Delta\chi^2(5) = 15.49$, $p = .002$; and summarization, $\Delta\chi^2(5) = 15.49$, $p = .002$. For recall, we could not simplify the full factorial model because almost all parameter estimates were significant. For the other outcome measures, we concluded that the simplified model had no reduced model fit when compared with the full factorial model: comprehension questions, $\Delta\chi^2(3) = 1.46$, $p = .69$; and summarization, $\Delta\chi^2(2) = 0.79$, $p = .67$. Subsequently, we added two instructional features to the model: teacher modeling (in addition to or instead of explicit instruction only) and individual student activities (in addition to or instead of collaborative activities only), both also in interaction with measurement timepoint. Adding these instructional components did not improve model fit for immediate effect sizes. For delayed effect sizes, adding individual practice led to a marginally significant⁵ improvement of model fit for comprehension questions, $\Delta\chi^2(4) = 9.39$, $p = .052$.

⁵ For the record, this effect is worth mentioning, even though it is not significant. A more powerful analysis in which the delayed effects on questions, recall, and summarization were collapsed and task type was modeled as an interaction effect, at this point showed a significant improvement in model fit ($\Delta\chi^2(2) = 6.84$, $p = .03$), as well as a significant parameter estimate for individual practice on delayed effect measures ($\Delta g = -0.74$, $SE = 0.31$, $p = .02$).

Model fitting procedures: Comprehension questions

Table A1 provides an overview of the model fitting procedure for the effect sizes related to comprehension questions. After considering the impact of methodological variables (M1, M2, M3), we examined text-related aspects (M4, M5), content-related moderating variables (M6), and instructional variables (M7). The model fit for reduced M6c is not significantly different from the full factorial M6b, $\Delta\chi^2(3) = 1.46$, $p = .69$; similarly, the model fit for reduced M7b is not significantly different from the full M7a, $\Delta\chi^2(1) = 1.83$, $p = .18$. The final parameter estimates are based on M7b. Table A2 presents the resulting parameter estimates for comprehension questions.

TABLE A1

Overview of stepwise model fitting for comprehension questions

Model	-2LL	Comparison	$\Delta\chi^2$	Δdf	p
M0	145.10				
M1: M0 + Δ Published	143.78	M0 vs. M1	1.32	1	.24
M2: M0 + Δ Competing_Intervention	143.92	M0 vs. M2	1.79	1	.28
M3: M0 + Δ Time	138.22	M0 vs. M3	1.79	1	.009
M4a: M3 + Δ Genre	138.20	M3 vs. M4a	0.12	1	.73
M4b: M3 + Δ Genre + (Δ Genre \times Time)	137.20	M3 vs. M4b	1.01	2	.60
M5a: M3 + Δ SN	137.67	M3 vs. M5a	0.54	1	.46
M5b: M3 + Δ SN + (Δ SN \times Time)	134.77	M3 vs. M5b	3.44	2	.18
M6a: M3 + Δ SV + Δ SS	137.48	M3 vs. M6a	0.73	2	.69
M6b: M6a + Δ SV_Active + Δ SS_Rule-Based + Δ Paragraph_Level	130.19	M3 vs. M6b	8.02	5	.15
M6c: M3 + Δ SV + Δ SV_Active	131.64	M3 vs. M6c	6.57	2	.04
M6d: M6c + (Δ SV \times Time) + (Δ SV_Active \times Time)	129.20	M6c vs. M6d	2.44	2	.29
M7a: M6c + Δ MO + (Δ MO \times Time) + Δ NI + (Δ NI \times Time)	122.26	M6c vs. M7a	9.39	4	.052
M7b: M6c + Δ MO + Δ NI + (Δ NI \times Timepoint)	122.25	M7a vs. M7b	1.83	1	.18

Note. MO = teacher modeling; NI = no individual activities; SN = structure number; SS = structure summarization; SV = structure visualization.

TABLE A2

Parameter estimates for the final model for comprehension questions

Model result	Estimate	SE	<i>P</i>
Intercept	0.98	0.30	.001
Delayed posttest	-1.21	0.37	.001
Structure-based visualization (overall)	-0.29	0.15	.06
Active structure mapping	0.39	0.15	.009
Teacher modeling	-0.09	0.10	.38
No individual activities: Posttest	0.45	0.23	.06
No individual activities: Delayed posttest	-1.04	0.39	.007

Model fitting procedures: Recall

Table A3 shows the model fitting procedure for recall. We successively examined methodological (M1, M2, M3), text-related (M4, M5), content-related (M6), and instructional variables (M7). The parameter estimates are based on M6b (Table A4).

TABLE A3

Overview of stepwise model fitting for recall

Model	-2LL	Comparison	$\Delta\chi^2$	Δdf	<i>p</i>
M0	133.03				
M1: M0 + Δ Published	128.03	M0 vs. M1	5.00	1	.025
M2: M0 + Δ Competing Intervention	132.59	M0 vs. M2	0.44	1	.51
M3: M0 + Δ Time	128.18	M0 vs. M3	4.84	1	.028
M4a: M3 + Δ Genre	128.08	M3 vs. M4a	0.11	1	.74
M4b: M3 + Δ Genre + (Δ Genre \times Time)	128.16	M3 vs. M4b	1.01	2	.90
M5a: M3 + Δ SN	128.18	M3 vs. M5a	0.003	1	.96
M5b: M3 + Δ SN + (Δ SN \times Time)	128.77	M3 vs. M5b	0.02	2	.99
M6a: M3 + Δ SV + Δ SS	128.10	M3 vs. M6a	0.08	2	.96
M6b: M6a + Δ SV_Active + Δ SS_Rule-Based + Δ Paragraph_Level	112.69	M3 vs. M6b	15.49	5	.002
M6c: M6b + Time (Δ SV_Active + Δ SS_Rule-Based + Δ Paragraph_Level)	109.66	M6b vs. M6c	3.03	4	.55
M7: M6b + Δ MO + (Δ MO \times Time) + Δ NI	110.18	M6b vs. M7	2.51	3	.47

Note. MO = modeling; NI = no individual activities; SN = structure number; SS = structure summarization; SV = structure visualization. At model M7, none of the interventions without individual activities administered a delayed posttest.

TABLE A4
Parameter estimates for the final model for recall.

Model result	Estimate	SE	<i>p</i>
Intercept	1.03	0.39	.009
Delayed posttest	-0.18	0.11	.09
Paragraph-level structure instruction	0.57	0.29	.03
Structure-based visualization (overall)	-0.44	0.18	.02
Active structure mapping	0.51	0.17	.002
Structure-based summarization (overall)	-0.24	0.17	.14
Rule-based summarization technique	0.34	0.12	.004

Model fitting procedures: Text structure knowledge

Table A5 provides an overview of the model fitting procedure for the effect sizes related to text structure knowledge. As there was a low number of observations, we could not estimate the moderating impact of most content-related or instructional variables. After considering the impact of methodological variables (M1, M2), we examined the impact of the number of different text structures taught (M3). The final parameter estimates are based on M3 and are shown in Table A6).

TABLE A5
Overview of stepwise model fitting for text structure knowledge

Model	-2LL	Comparison	$\Delta\chi^2$	Δdf	<i>p</i>
M0	12.16				
M1: M0 + Δ Published	12.39	M0 vs. M1	0.23	1	.63
M2: M0 + Δ Competing Intervention	13.30	M0 vs. M2	1.14	1	.29
M3: M0 + Δ Number of structures	16.88	M0 vs. M3	4.64	1	.03

TABLE A6
Parameter estimates for the final model for text structure knowledge

Model result	Estimate	SE	<i>p</i>
Intercept	0.57	0.11	<.001
Number of different structures taught	-0.06	0.02	.02

Model fitting procedures: Summarization

Table A7 provides the model fitting procedure for the summarization effect sizes. After considering the impact of methodological variables (M1, M2, M3), we examined text-related aspects (M4), content-related variables (M5), and instructional variables (M6). The model fit for reduced M5c is not significantly different from the full M5b, $\Delta\chi^2(2) = 0.79$, $p = .67$. For M6, there were no interventions with a delayed posttest without individual practice during the intervention. Therefore, final parameter estimates for summarization are based on M5c (see Table A8).

TABLE A7

Overview of stepwise model fitting for summarization

Model (M)	-2LL	Comparison	$\Delta\chi^2$	Δdf	p
M0	112.64				
M1: M0 + Δ Published	108.69	M0 vs. M1	3.95	1	.05
M2: M0 + Δ Competing Intervention	110.49	M0 vs. M2	2.16	1	.14
M3: M0 + Δ Time	102.01	M0 vs. M3	9.73	1	.002
M4a: M3 + Δ SN	102.70	M3 vs. M4a	0.21	1	.65
M4b: M3 + Δ SN + (Δ SN \times Time)	102.71	M3 vs. M4b	0.21	2	.90
M5a: M3 + Δ SV + Δ SS	97.13	M3 vs. M5a	5.78	2	.056
M5b: M5a + Δ SV_Active + Δ SS_Rule-Based + Δ Paragraph_Level	81.31	M3 vs. M5b	15.49	5	.002
M5c: M3 + Δ SS + Δ SS_Rule-Based + Δ Paragraph_Level	82.10	M5b vs. M5c	0.79	2	.67
M5d: M5c + (Time \times M5c)	80.78	M5c vs. M5d	1.32	3	.72
M6: M5c + Δ MO + (Δ MO \times Time) + Δ NI	79.15	M5c vs. M6	9.39	3	.40

Note. MO = teacher modeling; NI = no individual activities; SN = structure number; SS = structure summarization; SV = structure visualization.

TABLE A8

Parameter estimates for the final model for summarization

Model result	Estimate	SE	p
Intercept	0.22	0.40	.58
Delayed posttest	-0.47	0.16	.004
Paragraph-level structure instruction	0.91	0.22	<.001
Structure-based summarization (overall)	-0.14	0.26	.61
Rule-based summarization technique	0.64	0.21	.005

Appendix B: Mixed-methods analysis (Chapter 3)

This QR-code is linked to an elaborate overview of all aspects and/or questions related to the curricular structure and objectives, lesson goals, theory and instruction, student activities and transfer that guided the materials analysis, teachers interviews, and lesson observations.



Appendix C: Example lesson and text structure test (Chapter 5)

Example lesson

This QR-code is linked to a translated lesson example from the original TOP materials in the version that was tested during the intervention study. We share these materials to promote replicability and to provide insight into the exact content.



Text structure knowledge test

The next pages show several translated items from the text structure test to give an impression of the types of knowledge that were tested.

Example item 1: Text structure recognition

Read *Text 1*. What does the author mainly intend to convey in text 1?

- A The causes and consequences of something (cause-effect)
- ☒ B Differences and similarities (compare-contrast)
- C The order in which things (should) happen (chronology)
- D A problem and its solutions (problem-solution)

Text 1: Onions or shallots

Shallots and onions appear much alike, except a shallot has a finer taste than an onion. When you heat up a shallot, its sugars will start to caramelize. This does not happen when you heat up onions. Also there is a difference in how they grow: onions grow as one bulb, whereas shallots separate into clusters of smaller individual bulbs. Similar to onions, shallots can irritate the eyes when you peel and slice them. Both onions and shallots can be stored for up to two months in a cool, dark, and non humid place.

Example item 2: Explicit knowledge: cue words

Jayden reads a text about the causes of tornados. Which cue words will he probably find in this text? Circle the right answer.

- A One way to, because, a strategy for
- B After, when, eventually
- ☒ C Causes, thereby, because of, is produced by
- D Different, other than, the same

Example item 3. Explicit knowledge: questioning

Job reads a difficult text on how iron and steel are manufactured in a process of many different steps. During reading, he wants to check if he still understands what he reads. What question should he ask himself to monitor his comprehension?

- A Which words do I need to look up in the dictionary?
- B What are differences between iron and steel?
- C What effects are mentioned in the text?
- ☒ D What is the first step, what is the second step, ...?

Example item 4: Explicit knowledge: title and predictions

Look at mister Thom's book shelves. He wants to teach his students how to recognize causes and effects. Circle all the books that he can probably use for this.

Example item 5. Structure-based (short response) summary

Read *Text 2* and complete the summary.

Internet use can lead to various computer problems.

Problem 1 is: spam. The solution for this: a spam filter.

Problem 2 is: computer viruses. The solution for this: antivirus software.

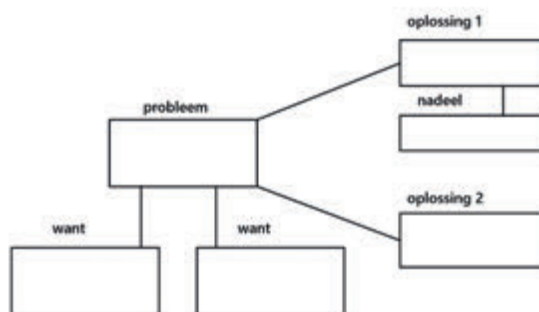
Text 2: Protecting your computer

Internet use involves all kinds of risks. For instance, spam is unsolicited messages sent for the purpose of advertising unwanted products or services. Another problem is computer viruses: small malware programs that are hidden in other files. Once they get into your computer, they do great damage. You can protect your computer from viruses by installing an antivirus software package. That way they cannot harm your computer. Spam is annoying but no serious threat. In order to receive less spam on your computer, you can use a spam filter in your email inbox.

Example item 6. Structure-based graphic-organizer

Use the information from Text 3 and put the numbers of the sentences listed below in the right place in the diagram. You will end up with two features that do not belong in any of the circles.

1. Use guard dogs
2. Comeback of the wolf
3. Build better stables
4. Wolves kill animals
5. Electric fences
6. Disease carriers
7. Stop them at the border
8. Jump high over fences



Text 3: Dutch prepare for wolf comeback

Until now, only lone wolves turned up in the Netherlands. Chances are increasing that wolves will soon come in numbers. The Netherlands is preparing for the comeback of the wolf, because we cannot stop them at the border. Of course, the arrival of the wolf creates problems: wolves are predators. They have a great preference for deer, sheep, calves and foals, but poultry is not safe either. In addition, wolves can transmit all sorts of dangerous diseases to livestock, as they often carry numerous parasites. Some of these diseases are dangerous to humans. Fortunately, there are a few ways to outsmart wolves. You can install electric fences everywhere to fence off pastures. These high fences with electrical voltage keep wolves out of the pasture. The downside to this is that wolves can learn to jump over them. Therefore, a better solution is the guard dog: guard dogs are well suited to guard herds of animals, especially at night. When two or three guard dogs are used, a wolf is not likely to attack an animal again. Even if the wolf jumps over a fence, guard dogs will chase it away quickly.

Appendix D: Model fitting procedures and estimates (Chapter 5)

Successive model fitting procedures

Table D1 presents the model fitting procedures for summarization, metacognitive knowledge, and writing. As the text structure knowledge test and standardized reading test were analyzed at item level with generalized linear models, so that the -2LL is no reliable measure for model comparison; instead, for these outcome measures, we used F-ratios of fixed effects to assess model fit (see Table D2).

TABLE D1 Successive model fitting for summarization, metacognition, and writing

	-2LL	Comparison			
		Models	$\Delta\chi^2$	Δdf	p
Summarization					
M0	5156.26	-	-	-	-
M1: M0 + Grade	5145.79	M0 vs M1	10.47	1	<.001
M2: M1 + Time	5120.12	M1 vs M2	25.67	2	<.001
M3: M2 + Group	5119.19	M2 vs M3	0.93	1	.33
M4: M3 + Time * Group	5100.74	M3 vs M4	18.45	2	<.001
M5: M4 + Grade * Time	5097.51	M4 vs M5	3.23	2	.20
M6: M5 + Grade * Group	5097.24	M5 vs M6	0.27	2	.87
M7: M6+ Grade * Time * Group	5094.62	M6 vs M7	2.62	2	.27
Metacognitive knowledge					
M0	2698.00	-	-	-	-
M1: M0 + Grade	2684.17	M0 vs M1	13.83	1	<.001
M2: M1 + Time	2661.98	M1 vs M2	22.19	2	<.001
M3: M2 + Group	2661.94	M2 vs M3	0.04	1	.84
M4: M3 + Time * Group	2651.27	M3 vs M4	10.67	2	.005
M5: M4 + Grade * Time	2647.97	M4 vs M5	3.30	2	.19
M6: M5 + Grade * Group	2645.17	M5 vs M6	2.80	2	.25
M7: M6 + Grade * Time * Group	2640.30	M6 vs M7	4.87	2	.09
Writing					
M0	1535.71	-	-	-	-
M1: M0 + Grade	1522.46	M0 vs M1	13.25	1	< .001
M2: M1 + Time	1517.58	M1 vs M2	4.88	2	.09
M3: M2 + Group	1515.91	M2 vs M3	1.67	1	.20
M4: M3 + Time * Group	1507.40	M3 vs M4	8.51	2	.02
M5: M4 + Grade * Time	1506.88	M4 vs M5	0.52	2	.77
M6: M5 + Grade * Group	1505.76	M5 vs M6	1.12	2	.57
M7: M6 + Grade * Time * Group	1504.26	M6 vs M7	1.50	2	.47

Random part: classes, students * classes, and error.

TABLE D2

Fixed effects in the final models per outcome measure

Text structure test	<i>F</i>	df ₁	df ₂	<i>p</i>
Grade	7.47	1	10556	.006
Time	3.40	2	10556	.03
Group	0.15	1	10556	.70
Time * Group	21.30	2	10556	< .001
Grade * Time	3.11	2	10556	.04
Grade * Group	0.32	1	10556	.57
Grade * Time * Group	4.62	2	10556	.01
Standardized reading test				
Grade	4.30	1	7418	.04
Time	0.33	2	7418	.72
Group	0.95	1	7418	.33
Time * Group	4.53	2	7418	.01
Grade * Time	9.81	2	7418	< .001
Grade * Group	0.68	1	7418	.41
Grade * Time * Group	3.34	2	7418	.04
Summarization				
Grade	14.12	1	19.8	.001
Time	13.90	2	553.3	< .001
Group	0.87	1	10.5	.37
Time * Group	9.38	2	553.1	< .001
Metacognitive knowledge				
Grade	21.61	1	21.6	.001
Time	12.47	2	12.5	< .001
Group	0.05	1	8.3	.83
Time * Group	5.39	2	571.7	.005
Writing				
Grade	31.48	1	9.4	.003
Time	2.39	2	561.8	.09
Group	1.57	1	8.4	.24
Time * Group	4.31	2	561.3	.014

Parameter estimates for summarization, metacognitive knowledge, and writing

Table D3 shows the estimated means and standard errors for summarization, metacognitive knowledge, and writing. The dashed line represents the switch in conditions: TOP-1 received the intervention between Time 1 and 2, TOP-2 between Time 2 and 3.

TABLE D3
Estimated means and standard errors per outcome measure

Summarization (scale 0-100)	TOP-1 Grade 4	TOP-2 Grade 4	TOP-1 Grade 5+	TOP-2 Grade 5+
Time 1	24.33 (4.1)	18.61 (4.1)	39.08 (4.1)	33.36 (4.1)
Time 2	38.15 (4.0)	25.00 (4.0)	52.91 (4.0)	39.76 (4.0)
Time 3	29.33 (4.1)	36.01 (4.0)	44.09 (4.1)	50.76 (4.1)
S^2 between classes	37.67 (29.9)			
S^2 students in classes	497.32 (20.6)			
Metacognitive knowledge (scale 0-15)				
Time 1	9.99 (0.35)	8.87 (0.38)	11.27 (0.32)	11.08 (0.33)
Time 2	11.03 (0.36)	10.42 (0.37)	11.46 (0.32)	11.76 (0.33)
Time 3	10.57 (0.36)	10.99 (0.37)	11.52 (0.32)	12.86 (0.33)
S^2 between classes	0.12 (0.32)			
S^2 students in classes	5.45 (0.32)			
Writing (scale 0-5)				
Time 1	1.94 (0.10)	2.04 (0.11)	2.43 (0.10)	2.53 (0.11)
Time 2	2.24 (0.10)	1.82 (0.11)	2.73 (0.10)	2.31 (0.11)
Time 3	2.19 (0.10)	2.17 (0.11)	2.68 (0.10)	2.66 (0.11)
S^2 between classes	0.81 (0.05)			
S^2 between students	0.005 (0.009)			

Parameter estimates for text structure test and the standardized reading test

Table D4 presents the estimates and standard deviations for the text structure test and the standardized reading test. As these estimates stem from a generalized linear model, the estimates should be interpreted as logits. For the figures in the running text of *Chapter 5*, these values were transformed in proportions per time, group, and grade.

TABLE D4

Estimates in logits for the text structure test and the standardized reading test

Model result	Text structure test		Standardized reading test	
	<i>Estimate</i>	<i>SE</i>	<i>Estimate</i>	<i>SE</i>
Grade 4	0.66	0.44	0.90	0.54
Grade 5+	0.61	0.45	0.55	0.53
Time 1	-1.46	0.56	-0.97	0.68
Time 2	-0.33	0.53	-1.19	0.73
TOP-1 group	-0.81	0.30	-0.77	0.29
Time 1 * TOP-1	0.88	0.19	0.58	0.20
Time 2 * TOP-2	1.02	0.17	0.79	0.22
Time 1 * Grade 5+	0.51	0.18	0.83	0.20
Time 2 * Grade 5+	0.54	0.17	0.82	0.22
TOP-1 * Grade 5+	0.72	0.40	0.70	0.38
TOP-1 * Grade 5+ * Time 1	-0.74	0.24	-0.51	0.27
TOP-1 * Grade 5+ * Time 2	-0.50	0.24	-0.74	0.30
S^2 between classes	0.048	0.05	0.05	0.04
S^2 students in classes	0.66	0.08	0.40	0.06
S^2 between items	2.33	0.53	2.89	0.80

Appendix E: Implementation fidelity measures

Lesson observations

The implementation of TOP lessons was observed by making use of the following rubric (Table E1). Observers were encouraged to score each of the numbered implementation aspects, take notes of lesson duration and student involvement, and to make extensive notes for all deviations from the manual.

TABLE E1
Rubric for the lesson observations

Lesson phase: <i>Explicit Instruction</i>	Implemented as intended?	Remarks
1. Teacher discusses lesson goals	<input type="radio"/>	
2. Teacher provides text structure instruction	<input type="radio"/>	
3. Teacher provides clear instruction	<input type="radio"/>	
4. Teacher reviews the previous lesson(s)	<input type="radio"/>	
Student involvement	0 –1 – 2	
Duration	___ minutes	
Overall impression:		
Lesson phase: <i>Modeling</i>	Implemented as intended?	Remarks
5. Teacher announces modeling	<input type="radio"/>	
6. Teacher gives concrete assignment	<input type="radio"/>	
7. Teacher uses I-pronoun for think aloud	<input type="radio"/>	
8. Teacher does not engage in interaction (i.e., no interactive questions <i>during</i> think aloud)	<input type="radio"/>	
9. Teacher does not simply read modeling script	<input type="radio"/>	
10. Teacher models with intended focus on text structure	<input type="radio"/>	
11. Teacher guides reflective discussion as intended by the teaching manual.	<input type="radio"/>	
Student involvement	0 –1 – 2	
Duration	___ minutes	
Overall impression:		

TABLE E1
Rubric for the lesson observations (continued)

Lesson phase: <i>Guided practice</i>	Implemented as intended?	Remarks
12. Teacher provides guided practice as intended by the teaching manual.	<input type="radio"/>	
<i>Check if activity is realized</i>		
A. Students read aloud.	<input type="radio"/>	
B. Students answer questions initiated by teacher.	<input type="radio"/>	
C. Several students get to answer the question.	<input type="radio"/>	
D. Students get to correct each other's answers.	<input type="radio"/>	
E. Teacher provides correct answers.	<input type="radio"/>	
F. Teacher provides additional explanation.	<input type="radio"/>	
Student involvement	0 – 1 – 2	
Duration	___ minutes	
Overall impression:		
Lesson phase: <i>Collaborative learning</i>	Implemented as intended?	Remarks
13. Teacher enacts collaborative task as intended by the teaching manual.	<input type="radio"/>	
14. Teacher gives instruction to promote collaborative learning.	<input type="radio"/>	
15. Teacher provides spontaneous/requested* support during collaborative learning task.	<input type="radio"/>	
16. Teacher provides feedback on the task afterwards.	<input type="radio"/>	
Student involvement	0 – 1 – 2	
Duration	___ minutes	
Overall impression		

TABLE E1

Rubric for the lesson observations (continued)

Observation of Student Pair 1/Student Pair 2 (during expert AAA groups/mixed ABC groups*).*

<i>Check if activity is realized:</i>	Yes / No	Remarks
A. Students show on-task behavior most of the time.	<input type="radio"/>	
B. Students execute task in mixed (ABC) groups.	<input type="radio"/>	
C. Students ask for teacher support/feedback.	<input type="radio"/>	
D. Students receive teacher support/feedback.	<input type="radio"/>	
E. Students finish task.	<input type="radio"/>	
<i>(This section is repeated 4 times at the original sheet, so that two student pairs could be observed during both parts of the collaborative learning task)</i>		
Lesson phase: <i>Individual practice</i>	Implemented as intended?	Remarks
17. Teacher enacts individual practice as intended by the manual.	<input type="radio"/>	
18. Teacher provides spontaneous/requested* support during individual practice.	<input type="radio"/>	
19. Teacher provides feedback on the task as intended.	<input type="radio"/>	
Student involvement	0 –1 – 2	
Duration	___ minutes	
Overall impression		
Lesson phase: <i>Lesson Closure</i>	Implemented as intended?	Remarks
20. Teacher enacts lesson closure as intended by the manual.	<input type="radio"/>	
Student involvement	0 –1 – 2	
Duration	___ minutes	
Overall impression		

TABLE E2
Main questions and follow-up questions for semi-structured interview

Main questions	Follow-up questions
1. Can you comment on the <i>content</i> of the lesson series? How did you experience this?	Did you experience any challenging issues with explaining text structure? Were you already familiar with the concept of text structure before teaching these lessons? How do you evaluate the degree of support from the manual and/or teacher video clips on text structure?
2. How satisfied are you with the <i>texts</i> in the lesson series?	How suitable do you feel these texts were for your average students? How suitable were these texts for low-achieving students, and for high-achieving students?
3. Can you comment on the prescribed <i>teacher-centered activities</i> in the lesson series (explicit instruction, modeling, and guided practice)?	Were these lesson phases clear for you and your students? Did you intentionally do things differently for these activities than intended by the manual? Why?
4. Can you comment on the prescribed <i>student-centered activities</i> in the lesson series (collaborative learning, independent practice)?	Were these lesson phases clear for you and your students? Did you intentionally do things differently for these activities than intended by the manual? Why?
5. Can you comment on the content and pedagogy of this lesson series in relation to what you usually do during reading comprehension lessons?	To what extent was this lesson series different or similar? How satisfied were you with the lesson series overall?

TABLE E3

Translated version of logbooks that were used during the multiple-case study
(Chapter 6)

Name:	Lesson number:																																
Did you split the lesson in two parts?	0 No 0 Yes, after phase _____																																
How satisfied are you with the lesson duration? (0: too short, 5: too long)	Rate: 0 - 5																																
How did you implement each of the following lesson parts:	<table><tr><td><i>As intended</i></td><td><i>Longer</i></td><td><i>Shorter</i></td><td><i>Omitted</i></td></tr><tr><td>1. Explicit instruction</td><td>0</td><td>0</td><td>0</td></tr><tr><td>2. Modeling</td><td>0</td><td>0</td><td>0</td></tr><tr><td>3. Reflection after modeling</td><td>0</td><td>0</td><td>0</td></tr><tr><td>4. Guided practice</td><td>0</td><td>0</td><td>0</td></tr><tr><td>5. Collaborative activity</td><td>0</td><td>0</td><td>0</td></tr><tr><td>6. Individual practice</td><td>0</td><td>0</td><td>0</td></tr><tr><td>7. Lesson closure</td><td>0</td><td>0</td><td>0</td></tr></table>	<i>As intended</i>	<i>Longer</i>	<i>Shorter</i>	<i>Omitted</i>	1. Explicit instruction	0	0	0	2. Modeling	0	0	0	3. Reflection after modeling	0	0	0	4. Guided practice	0	0	0	5. Collaborative activity	0	0	0	6. Individual practice	0	0	0	7. Lesson closure	0	0	0
<i>As intended</i>	<i>Longer</i>	<i>Shorter</i>	<i>Omitted</i>																														
1. Explicit instruction	0	0	0																														
2. Modeling	0	0	0																														
3. Reflection after modeling	0	0	0																														
4. Guided practice	0	0	0																														
5. Collaborative activity	0	0	0																														
6. Individual practice	0	0	0																														
7. Lesson closure	0	0	0																														
How satisfied are you with each of the following lesson parts:	Rate: 0 – 5																																
1. Explicit instruction																																	
2. Modeling and refection																																	
3. Guided practice																																	
4. Collaborative learning																																	
5. Individual practice																																	
Please explain your answer with regard to lesson parts 1 to 3.	Open response																																
Please explain your answer with regard to lesson parts 4 and 5.	Open response																																
How satisfied are you with the texts in this lesson?	Rate: 0 – 10																																
How suitable was this lesson for your students?	Rate: 0 – 10																																
What is your overall rating of the lesson?	Rate: 0 – 10																																
How could this lesson be improved?	Open response																																

SAMENVATTING, DANKWOORD & about the author



Samenvatting

1. Wat is tekststructuur?

Een tekst getiteld *Corona of griep?* heeft waarschijnlijk een andere onderliggende tekststructuur dan de tekst *Corona: van eerste ontdekking tot pandemie*. Weliswaar gaan beide teksten over corona, maar de eerste tekst stelt vooral verschillen en overeenkomsten centraal, terwijl de tweede tekst de focus legt op de stapsgewijze verspreiding van het virus. Met andere woorden: de eerste tekst heeft een vergelijkingsstructuur, de tweede een chronologische structuur. Een tekststructuur is als het ware het skelet van de tekst waaraan de informatie is opgehangen. Het betreft de organisatie van ideeën in een tekst, de relaties tussen die ideeën, en de talige middelen die gebruikt worden om die relaties expliciet te maken (Pyle et al., 2017). Net zoals giraffen en dinosaurussen een verschillend skelet hebben, hebben teksten dit ook. Tabel 1 toont voorbeelden van de vier tekststructuren die centraal stonden in dit proefschrift: de vergelijking, de oorzaak-gevolgstructuur, de chronologische structuur en de probleem-oplossingsstructuur (zie Meyer, 1975).

Goede lezers proberen tekstinhoud zo goed mogelijk te ordenen om zo de rode draad van de tekst te zien. Hiervoor moeten ze hoofdgedachtes kunnen filteren en informatie uit verschillende tekstdelen aan elkaar verbinden. Daarbij kan onderwijs over tekststructuur helpen: als je leerlingen leert welke tekststructuren er bestaan, worden ze beter in het herkennen van hoofdgedachtes en in het integreren van informatie (vgl. Williams, 2018).

Precies deze vaardigheid tot integreren lijkt bij Nederlandse leerlingen ondermaats volgens peilingonderzoeken (PEIL.TAAL-2019; Onderwijsinspectie, 2020; PIRLS-2016; Gubbels et al., 2017; PISA-2018; Gubbels et al., 2019). Dat is problematisch, omdat in de bovenbouw van de basisschool de hoeveelheid informatieve teksten fors toeneemt en leerlingen steeds meer moeten gaan lezen-om-te-leren (Chall & Jacobs, 1983). Daarom onderzocht ik in dit proefschrift de mogelijkheden voor en effecten van expliciet tekststructuuronderwijs in de bovenbouw van de basisschool. Voordat ik de opzet en uitkomsten van mijn onderzoek bespreek, leg ik eerst uit welke rol tekststructuur kan spelen bij de totstandkoming van tekstbegrip.

TABEL 1

Vier tekststructuren met gecursiveerd enkele kenmerkende signaalwoorden

<p><i>Vergelijkingsstructuur</i> Tekst vergelijkt twee of meer zaken of gebeurtenissen</p>	<p><i>Oorzaak-gevolgstructuur</i> Tekst verklaart hoe de ene gebeurtenis voortvloeit uit een andere</p>
<p>CORONA OF GRIEP?</p> <p>Griep en corona zijn <i>allebei</i> besmettelijke virusinfecties die zich vaak nestelen in de luchtwegen, <i>maar</i> ze worden veroorzaakt door <i>verschillende</i> virussen. Griep wordt veroorzaakt door het influenzavirus, en COVID-19 door een coronavirus. De ziektebeelden <i>lijken</i> veel op elkaar. <i>Beide</i> virussen veroorzaken koorts, moeheid en kortademigheid en gaan vaak gepaard met hoesten en keelpijn.</p>	<p>WAARDOOR ONTSTAAN</p> <p>BLOEDKLONTERS BIJ COVID-19?</p> <p>Wetenschappers weten nog steeds niet precies <i>waardoor</i> COVID-19 voor bloedklonters <i>zorgt</i>. Eén mogelijke theorie is dat deze <i>ontstaan</i> als een <i>reactie van</i> bloedvaten op het virus. Het virus bindt zich aan ACE2-receptoren die in bloedvaten te vinden zijn. <i>Als</i> het coronavirus zich bindt, <i>dan</i> ontsteken de vaten, en dat <i>veroorzaakt</i> bloedklonters.</p>
<p><i>Chronologische structuur</i> Tekst bespreekt tijdsvolgorde of stappen in een proces.</p>	<p><i>Probleem-oplossingsstructuur</i> Tekst noemt een probleem en de mogelijke oplossing(en).</p>
<p>VAN ONTDEKKING TOT PANDEMIE</p> <p><i>Op 1 januari 2020</i> werd de markt in Wuhan gesloten voor inspectie en desinfectie, omdat het er de schijn van had dat veel mensen daar besmet waren geraakt. <i>Op 7 januari</i> stelden Chinese gezondheidsautoriteiten vast dat het ging om een nieuw virus. <i>Vijf dagen later</i> deelde China de genetische code van het nieuwe virus met de wereld. <i>Niet veel later</i> werd het virus wereldwijd aangetroffen en was de eerste coronapandemie een feit.</p>	<p>MONDKAPJESPRODUCTIE OP DE RAILS</p> <p>Door de coronacrisis ontstond een wereldwijd tekort aan mondkapjes. Er raakten ook veel mensen werkloos door lockdowns. <i>Beide problemen</i> werden creatief <i>aangepakt</i> in India: werkloos geraakt treinpersoneel werd ingezet in fabrieken om mondkapjes te maken voor de zorg. <i>Op die manier</i> hadden deze mensen weer inkomen én werd het tekort aan mondkapjes <i>opgelost</i>.</p>

2. Welke rol speelt tekststructuur bij tekstbegrip?

Begrijpend kunnen lezen veronderstelt onder andere een rijke woordenschat, relevante voorkennis en vlotte decodeervaardigheden. Dat alleen is echter niet voldoende. Zodra teksten moeilijker worden en leerlingen vaker tegen begripsproblemen aanlopen, of wanneer leerlingen meer gaan lezen-om-te-leren, moeten ze leesstrategieën flexibel kunnen inzetten: op de juiste manier en op het juiste moment (Afflerbach & Cho, 2009; Cromley & Azevedo, 2007).

Niet voor niets richt het Nederlandse leesonderwijs zich vaak op strategie-onderwijs, al zijn er kanttekeningen te plaatsen bij de invulling daarvan. Zo staat het beantwoorden van vragen bij een tekst op de voorgrond tijdens lessen begrijpend lezen (Berends, 2011). Dit is een aanpak die steeds meer onder vuur ligt (Rooijackers et al., 2020; 2021), onder andere omdat deze leerlingen onvoldoende helpt om zelfregulerende, strategische lezers te worden (vgl. Afflerbach & Cho, 2009; Alexander, 2018; Paris et al., 1983).

Bovendien is het maar de vraag of strategie-onderwijs wel kan bestaan zonder aandacht voor tekststructuur. Lezers die zich bewust zijn van de tekststructuur kunnen die kennis inzetten om de belangrijkste informatie-eenheden uit de tekst te filteren en die op een logische manier met elkaar in samenhang te brengen (Meyer et al., 1980; Meyer & Freedle, 1984). De tekststructuur biedt als het ware een mentale kapstok om de tekstinhoud op een geordende manier aan op te hangen (bijvoorbeeld: de tijd als kapstok bij een chronologische tekststructuur). Interventieonderzoek – in vooral de Amerikaanse context – bevestigt dit: onderwijs over tekststructuur helpt leerlingen om tekstinhoud beter te begrijpen en onthouden (Hebert et al., 2016; Pyle et al., 2017).

Kennis over leesstrategieën en tekststructuur kunnen elkaar wederzijds versterken, omdat inzicht in de tekststructuur een gecontextualiseerd strategiegebruik kan bevorderen waarbij lezers leesstrategieën afstemmen op de specifieke tekststructuur (Hoch & McNally, 2020; Reutzel et al., 2005; Stevens & Vaughn, 2021). Tabel 2 geeft hiervan een voorbeeld: waar de aanwijzingen voor algemeen strategiegebruik abstract zijn (bijv. *Stel jezelf vragen tijdens het lezen*), geeft een structuurspecifiek strategiegebruik lezers juist concrete handvatten. Daarmee wordt het strategiegebruik minder afhankelijk van de metacognitieve vaardigheden waarover leerlingen al dan niet beschikken (Williams, 1998).

TABEL 2

Voorbeeld van algemeen versus structuurspecifiek strategiegebruik

Algemene strategie: vragen stellen	Strategie: vragen stellen bij een vergelijking
Stel jezelf vragen tijdens het lezen van de tekst. Dit helpt om je aandacht te richten op de tekst.	Deze tekst gaat over overeenkomsten en verschillen, dus je kunt jezelf afvragen: 1. Op welk kenmerk worden x en y vergeleken? 2. Betreft dit een verschil of een overeenkomst?
<i>Resultaat:</i> Vragen zijn algemene WH-vragen of moeten zelf bedacht worden, wat denkkraft kost en ook kan zorgen voor irrelevante focus.	<i>Resultaat:</i> Een specifieke vraagstelling passend bij tekstinhoud en structuur. Vragen hebben relevante focus en betreffen kerninformatie.

Deze visie sluit aan bij twee invloedrijke leestheorieën: het *Constructie-Integratiemodel* (Kintsch, 1988; 2004; 2013) en het *Landscape Model* (van den Broek et al., 1999). Beide modellen benadrukken dat teksten beter begrepen worden wanneer lezers steeds actief verbanden leggen tussen de informatie-eenheden binnen een tekst en tussen tekstinhoud en voorkennis. Tekststructuur helpt lezers om te zien waar de belangrijkste verbanden liggen tussen zinnen en alinea's en speelt daarom een belangrijke rol bij *top-down*-informatieverwerking, terwijl het actief aan de slag gaan met de tekst en het flexibel oplossen van begripsproblemen een vorm van *bottom-up*-informatieverwerking is. Beide processen helpen lezers om zich een samenhangende mentale representatie van de tekst te vormen.

3. Onderzoeksdoelen en bevindingen

Leesstrategieën hebben hun weg gevonden naar het klaslokaal, maar kennis over tekststructuur lijkt niet erg hoog op de agenda te staan van Nederlandse basisscholen en pabo's (Kooiker-den Boer et al., 2019; Scheltinga et al., 2013). Het is natuurlijk de vraag of dit terecht is. In de eerste fase van dit onderzoek heb ik

daarom een meta-analyse uitgevoerd (*Hoofdstuk 2*) waarmee ik onderzocht welke effecten tekststructuuronderwijs heeft op de leesvaardigheid van leerlingen in de bovenbouw van het primair onderwijs. Daarbij analyseerde ik ook welke inhoudelijke en didactische elementen gerelateerd zijn aan interventiesucces.

Mede dankzij de vrijheid van het onderwijs hebben educatieve uitgevers en scholen in Nederland veel ruimte om zelf het curriculum vorm te geven (Bruggink & Netten, 2017). De keerzijde hiervan is dat het relatief onduidelijk is op welke manier er aandacht wordt besteed aan tekststructuur en leesstrategieën en welke didactiek er gehanteerd wordt. In de tweede fase stond daarom de huidige praktijk centraal: wat bieden lesmaterialen en leerkrachten aan bij begrijpend lezen in groep 6 en 7?

In de derde fase heb ik geprobeerd om met de informatie uit de voorgaande hoofdstukken een brug te bouwen tussen wetenschap en praktijk. Eerst werden er ontwerpprincipes opgesteld die door een viertal leerkrachten werden gebruikt om lessen over tekststructuur te ontwerpen. Met een kwalitatieve insteek evalueerde ik in *Hoofdstuk 4* in hoeverre deze vier ontwerpprincipes toepasbaar waren in de praktijk en welke kennis leerkrachten nodig hebben om tekststructuuronderwijs te ontwerpen en te geven.

Een deel van de ontworpen lessen is vervolgens doorontwikkeld voor een effectstudie (*Hoofdstuk 5*) waarin de effecten op diverse leestoetsen, schrijfvaardigheid en expliciete metacognitieve kennis werden gemeten binnen een zogenaamd *switching-panels replication design* met tien klassen. Omdat de effecten bleken te verschillen per interventiegroep en er enige problemen leken te zijn met didactische aspecten, heb ik een aanvullende multiple-casestudy uitgevoerd (*Hoofdstuk 6*) om meer inzicht te krijgen in de mogelijke knelpunten bij de implementatie, met name bij de didactische aanpak.

3.1 Meta-analyse: heeft tekststructuuronderwijs eigenlijk wel zin?

De meta-analyse in *Hoofdstuk 2* bundelt de resultaten van 44 (quasi-) experimentele studies over tekststructuur, gericht op leerlingen in groep 6 tot en met 8. De eerste onderzoeksvraag was: wat zijn de directe en uitgestelde effecten van tekststructuur-onderwijs op begripsvragen, onthoudtaken, samenvattingen en tekststructuurkennis? Het bleek dat leerlingen na expliciet tekststructuuronderwijs beter presteerden op leestaken dan leerlingen die

traditioneel leesonderwijs of alternatieve interventies volgden. Wel verschilden de effectgroottes per type uitkomstmaat: $g = 0.25$ op tekstbegripsvragen, $g = 0.37$ op geheugentaken, $g = 0.57$ op samenvatten en $g = 0.38$ op tekststructuurkennis. Tekststructuuronderwijs bleek vooral effectief op korte termijn: op uitgestelde nametingen verdween het verschil met de controlegroep. Dit onderstreept hoe belangrijk het is om tekststructuuronderwijs duurzaam te verankeren in het curriculum van het primair onderwijs.

De tweede onderzoeksvraag was: welke inhoudelijke en didactische elementen beïnvloeden de effectiviteit van tekststructuuronderwijs? Een moderatoranalyse liet zien dat sommige elementen de effecten van tekststructuuronderwijs kunnen versterken. Dit geldt bijvoorbeeld voor het geven van uitleg over tekststructuur op alinea-niveau (zoals kernzinnen en details onderscheiden), het actief werken met structuurspecifieke *graphic organizers*, en het volgen van een structuurgebaseerd stappenplan bij het maken van een samenvatting. Een bevinding hierbij was ook dat het per uitkomstmaat varieert welke elementen effect sorteren: aandacht voor alineastructuur is bijvoorbeeld belangrijk voor samenvattingsvaardigheid, maar voegt niets extra toe als je een standaard leestoets geeft.

Bij de didactische aspecten bleek het niet uit te maken of een interventie ook modeling of samenwerkend leren bevatte. Wel bleek het voor langetermijneffecten van belang dat leerlingen ook zelfstandig geoefend hebben: zonder zelfstandige oefening gingen leerlingen namelijk sterker achteruit op de nameting dan wanneer ze de stof wel individueel hadden geoefend, ongeacht of ze dit daarvoor ook in groepjes of tweetallen hadden gedaan. Een belangrijke kanttekening bij deze uitkomst is overigens de nogal summiere beschrijving van didactische aspecten bij veel van de geïnccludeerde interventiestudies.

3.2 Wat leren Nederlandse leerlingen over tekststructuur in groep 6 en 7?

De precieze inhoud en didactiek van het Nederlandse curriculum voor begrijpend lezen zijn mede door de grondwettelijke onderwijsvrijheid relatief onduidelijk. Er zijn alleen einddoelen geformuleerd (het beoogde curriculum), maar scholen en educatieve uitgeverij kunnen zelf beslissen welke precieze inhoud en didactiek ze hanteren in de lesmaterialen (het geïmplementeerde curriculum) en hoe ze dit

toepassen in het klaslokaal (het uitgevoerde curriculum). In *Hoofdstuk 3* bracht ik daarom met een materiaalanalyse, interviews en lesobservaties het curriculum van begrijpend lezen in groep 6 en 7 in kaart.

De eerste onderzoeksvraag was: in hoeverre realiseren lesmethodes in hun inhoud een gebalanceerd aanbod over leesstrategieën en tekststructuren en hoe gebruiken en evalueren leerkracht deze inhoud van hun lesmethodes? Vanuit wetenschappelijke theorieën zou onderwijs over leesstrategieën en tekststructuur vooral effectief zijn wanneer leerlingen declaratieve kennis (het *wat*), procedurele kennis (het *hoe*) en conditionele kennis (het *wanneer* en *waarom*) kunnen opdoen (Paris et al. 1983). De analyse onthulde echter dat er momenteel geen gebalanceerd aanbod is in deze drie soorten kennis. De nadruk ligt op procedurele kennis, met een sterke focus op het oefenen van strategieën – vaak als een doel op zich, in plaats van een middel tot een doel.

Deze procedurele kennis staat vaak los van declaratieve kennis over tekststructuur: de uitleg hierover is zo minimaal – vaak beperkt tot lijstjes signaalwoorden zonder heldere toelichting over wat je ermee kunt – dat deze kennis nauwelijks zal bijdragen aan de ontwikkeling van hogere-orde-begripsvaardigheden. Interviews en lesobservaties lieten zien dat leerkrachten nauwelijks compenseren voor dit tekort in de lesmaterialen, al constateren zij wel problemen op dit vlak. Conditionele kennis krijgt eveneens nauwelijks aandacht: leerlingen hoeven bijna nooit hun eigen strategiegebruik te plannen of evalueren, en de aandacht voor transfer is beperkt. Het is maar de vraag of leerlingen voldoende leren om hun strategiegebruik te kunnen afstemmen op een leestaak.

Daarnaast bleek er in veel lesmethodes een matige afstemming te zijn tussen lesdoelen, uitleg en oefeningen, waardoor de focus van lessen vaak meer verschuift naar het tekstonderwerp, dan dat deze ligt op de beoogde lesdoelen – iets wat ook in de interviews naar voren kwam. Deze bevindingen sluiten naadloos aan bij die uit internationaal onderzoek: er is vaak een sterke nadruk op tekstinhoud tijdens leeslessen en veel minder aandacht voor tekststructuren en activiteiten gericht op dieper tekstbegrip (vgl. Wijekumar et al., 2021). Desondanks leunen leerkrachten sterk op hun lesmethodes, zelfs als ze de leerlijn daarin onduidelijk vinden.

De tweede onderzoeksvraag was gericht op de didactische aanpak: in hoeverre realiseren lesmethodes een *Gradual Release of Responsibility* (GRR;

Fisher & Frey, 2021; Pearson & Gallagher, 1983) en hoe gebruiken en evalueren leerkrachten de didactiek van hun lesmethode? Deze analyse toonde dat leerkrachten op dit vlak meer eigen accenten durven leggen. Toch is de impact van lesmaterialen groot: de nadruk op individueel vragen beantwoorden wordt verschillend ervaren, maar leerkrachten houden desondanks vaak vast aan deze werkvorm: bijna 43% van de geobserveerde lestijd bestond uit zelfstandig vragen beantwoorden. Het GRR-model komt onvoldoende uit de verf in de voorgeschreven en uitgevoerde lessen, mede doordat strategieën niet uitgebreid worden toegelicht en samenwerkend leren weinig ruimte krijgt. Lesmaterialen leggen sterke nadruk op *modeling* van strategieën, maar in de praktijk loopt dit vaak anders dan bedoeld: modeling wordt soms overgeslagen of ingekort, of wordt veranderd in interactieve instructie, waardoor het niet meer het karakter van observerend leren heeft.

3.3 Samen met leerkrachten een lessenserie over tekststructuur ontwerpen

Omdat tekststructuur nauwelijks op coherente wijze aan bod komt in Nederlandse lesmaterialen, bestond de volgende stap uit een ontwerpgericht onderzoek (zie Broekkamp & van Hout-Wolters, 2007; McKenney & Reeves, 2018) waarbij twee onderzoekers en vier leerkrachten samenwerkten bij het ontwikkelen van lessen over tekststructuur. Figuur 1 toont de vier principes die de basis vormden voor het lesontwerp. In twee cycli werden nieuwe lessen over tekststructuur ontworpen, getest, geëvalueerd en bijgesteld. Daarbij lag de focus niet alleen op productontwikkeling, maar vooral op het ontwerpproces. Met diverse databronnen (lesmaterialen, lesobservaties, logboeken, panelinterviews) wilde ik antwoord krijgen op de onderzoeksvraag: hoe haalbaar zijn de ontwerpprincipes in de praktijk en welke ondersteuning hebben leerkrachten nodig bij de toepassing ervan?

De toepassing van de ontwerpprincipes bleek allerm minst eenvoudig. Ten eerste belemmerde een matige kwaliteit van zaakvakteksten – vooral wat betreft hun structuur – de toepassing van het eerste ontwerpprincipe. Sommige leerkrachten pasten de zaakvakteksten aan op zinsniveau, maar slaagden er desondanks niet in om de structuur op alinea-niveau te verhelderen. Anderen lieten teksten zo authentiek mogelijk, maar ook dat ondermijnde de focus op

tekststructuur, omdat er dan ook veel leestijd werd besteed aan het complexe vocabulaire in deze teksten.

FIGUUR 1

Vier ontwerpprincipes voor de lessenserie over tekststructuur



Een andere belemmering – met name bij toepassing van het tweede en derde ontwerpprincipe – was het feit dat leerkrachten weinig kennis hadden over concepten zoals tekststructuur en conditionele kennis. Ze bespraken bijvoorbeeld wel de kenmerken van elke tekststructuur, maar worstelden met een structuurspecifieke toepassing van leesstrategieën, of neigden ernaar om de *tekstinhoud* sterker te benadrukken dan de *tekststructuur*. Op dit vlak was veel ondersteuning nodig van de onderzoekers, zodat leerkrachten hun vakdidactische kennis konden verdiepen.

Een derde aandachtspunt vormde de specifieke overtuigingen van sommige leerkrachten, bijvoorbeeld rondom teksten (authentieke of geredigeerde teksten) en modeling (perfect of lerend model). Dankzij de gedachtewisselingen in het ontwerpteam en het herhaaldelijk testen, evalueren en aanpassen van de materialen, kregen de leerkrachten meer grip op de ontwerpprincipes en herevalueerden ze soms overtuigingen die hiermee op gespannen voet stonden.

Het ontwerp onderzoek zorgde voor praktisch toepasbare kennis over de ontwerpprincipes en empirisch onderbouwde lesmaterialen, en droeg bij aan de professionalisering van de deelnemende leerkrachten. Tegelijkertijd onderstrepen de hobbels en kuilen van het ontwerp onderzoek dat *co-design* bij begrijpend lezen allerm minst eenvoudig is. Voor een duurzame brug tussen wetenschap en praktijk moet de vakdidactische kennis van leerkrachten worden vergroot (zie ook Kooiker-den Boer et al., 2019) en is het wenselijk dat educatieve uitgevers meer

investeren in de tekstkwaliteit van schoolboeken, vooral op het punt van tekststructuur.

3.4 Welk effect heeft de ontworpen lessenserie?

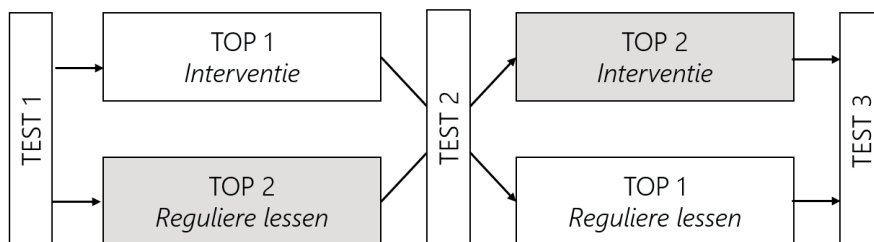
Een tiental lessen uit de ontwerpstudie werd doorontwikkeld tot de lessenserie TOP waarin leerlingen vier tekststructuren leerden herkennen: vergelijking, probleem-oplossing, chronologie en oorzaak-gevolg. Bovendien was er uitgebreid aandacht voor de manier waarop de leesstrategieën voorspellen, vragen stellen en samenvatten konden worden toegepast op een structuurspecifieke manier. De didactische opzet volgde het GRR-model waarbij expliciete instructie werd gevolgd door modeling en reflectie, gevolgd door een fase van begeleide inoefening en samenwerkend leren en tot slot een individuele (schrijf)opdracht. De groepsopdracht was geïnspireerd op een jigsawtaak (zie Aronson, 1978) waarin leerlingen in kleine groepjes een stuk tekst leerden samenvatten in graphic organizers en dit vervolgens aanvulden met informatie van medeleerlingen.

Hoofdstuk 5 beschrijft de effecten van de lessenserie op diverse uitkomstmaten: een tekststructuurtoets, een gestandaardiseerde leestoets, een samenvattingstaak, een vragenlijst over expliciete metacognitieve kennis en een schrijftaak. In totaal deden tien leerkrachten van zeven basisscholen mee samen met hun leerlingen ($N = 201$). Het onderzoek was opgezet volgens een zogenaamd *switching-panels replication design* met twee interventiegroepen (TOP 1 en TOP 2) en drie meetmomenten (Figuur 2; Shadish et al., 2002).

Na het eerste meetmoment volgde de ene helft van de klassen de interventie, terwijl de andere helft hun reguliere lessen volgde en zo als controlegroep fungeerde; na het tweede meetmoment wisselden de groepen om. Dit maakt het mogelijk om zowel de repliceerbaarheid van effecten als de langetermijneffecten in de eerste groep te onderzoeken.

FIGUUR 2

Switching-panels replication design met de interventiegroepen



De uitkomsten van het effectonderzoek waren minder consistent dan verwacht: de meeste positieve uitkomsten werden gevonden bij de interventiegroep die als tweede aan de beurt was (TOP 2). Leerlingen in deze groep gingen aantoonbaar vooruit in hun samenvattingsvaardigheid ($d = 0.48$) en expliciete metacognitieve kennis over leesstrategieën ($d = 0.24$). Alleen de leerlingen uit groep 6 in deze interventiegroep presteerden ook aantoonbaar beter op de gestandaardiseerde leestoets ($d = 0.53$) en de tekststructuurtoets ($d = 0.50$). Na afronding van het interventieprogramma scoorden alle leerlingen in beide interventiegroepen beter op de schrijftaak ($d = 0.33$ en $d = 0.39$).

Hoewel de resultaten erop lijken te wijzen dat kennis over tekststructuur leerlingen een handvat kan bieden om hoofdgedachtes te onderscheiden van details en om de juiste verbanden te leggen tussen informatie-eenheden binnen een tekst (Meyer et al., 1980; Miyatsu et al., 2018; Stevens & Vaughn, 2021; Strong, 2020) en om ook bij schrijftaken de tekstinhoud beter te organiseren (Dickson, 1999; Hebert et al., 2018), zijn de meeste effecten niet generaliseerbaar omdat ze maar in één iteratie bij een selecte groep leerlingen zijn gevonden.

Hiervoor zijn verschillende verklaringen mogelijk. Zo bleek de kwaliteit en paralleliteit van de gebruikte toetsen tegen te vallen – ondanks piloten en ijken – en sloten niet alle uitkomstmaten goed aan op de interventie (vgl. Clemens & Fuchs, 2021). Daarnaast waren er vragen rondom de implementatiekwaliteit, met name op het gebied van de didactiek. Uit de beperkte implementatiedata die beschikbaar waren, bleek dat de vorm van samenwerkend leren vaak werd aangepast. Ook werden individuele oefening en feedback ingekort, of zelfs overgeslagen door meerdere leerkrachten. Dit ondermijnde mogelijk het

onderliggende GRR-principe, wat problematisch kan zijn geweest gezien de relatief korte duur van de interventie in combinatie met de complexe inhoud.

3.5 Hoe implementeren leerkrachten de lessenserie?

Hoofdstuk 6 beschrijft de uitkomsten van een multiple-casestudy waarbij drie basisschoolleerkrachten intensief gevolgd zijn bij de implementatie van de enigszins herziene TOP-lessenserie. De data bestonden uit logboeken, herhaaldelijke lesobservaties ($N = 14$) en uitgebreide leerkrachtinterviews. Daarnaast werden op leerlingniveau observaties uitgevoerd tijdens de fase van samenwerkend leren en vulden leerlingen een vragenlijst in over hun ervaringen met de lessenserie. Het doel was om in kaart te brengen in hoeverre de aanpassingen die leerkrachten maakten in het interventiemateriaal de beoogde inhoud en didactiek (volgens het GRR-model) stimuleerden of juist hinderden.

Een veelvoorkomend probleem bij de implementatie van de GRR is dat leerkrachten weinig aandacht besteden aan de fases tussen expliciete instructie en individuele verwerking (Fisher & Frey, 2021). Dit bleek in deze multiple-casestudy niet het geval te zijn: leerkrachten besteedden juist relatief veel tijd aan de fases van begeleide inoefening en samenwerkend leren. Wel varieerden de implementatiescores per fase van het GRR-model: expliciete instructie (91%), modeling (62%), begeleide inoefening (58%), samenwerkend leren (50%) en individuele verwerking en lesafsluiting (50%). Elke leerkracht legde eigen accenten binnen het GRR-model: de een blonk uit in het geven van instructie en modeling, maar gaf weinig begeleiding tijdens de leerlingactiviteiten, terwijl een ander juist heel bedreven was in het stimuleren van samenwerkend leren, maar worstelde met instructie en modeling. Leerlingen en leerkrachten waren over het algemeen tevreden met de lessenserie, waarbij met name het samenwerkend leren en de functionele taken positief werden beoordeeld.

Natuurlijk zou een te strikte, rigide implementatie van het lesmateriaal onwenselijk zijn: leerkrachten moeten immers altijd vertaalslagen maken in het lesmateriaal om te zorgen dat dit optimaal aansluit bij de behoeften van hun klas op dat moment. Deze aanpassingen mogen echter nooit de onderliggende pedagogische principes ondermijnen. Hoewel er ook succesvolle aanpassingen waren in het lesmateriaal door de leerkrachten, bleken er in deze studie eveneens

implementatieproblemen te zijn die zich bij alle leerkrachten in meerdere lessen herhaalden, bijvoorbeeld bij het stimuleren van metacognitieve kennis en het begeleiden van samenwerkend leren.

Zo bleken leerkrachten de lesonderdelen gericht op metacognitieve kennis stelselmatig over te slaan: ze reflecteerden niet op hun leesproces na het modelen en sloegen vaak de transfer-/reflectievraag bij de lesafsluiting over, terwijl deze activiteiten juist een gesprek moesten initiëren over toekomstige situaties waarin het geleerde zinvol zou kunnen zijn. De begeleiding van het samenwerkend leren was voor twee leerkrachten eveneens ingewikkeld. Ze maakten op eigen initiatief aanpassingen in de opdrachten die het lesonderdeel weliswaar vlotter deden verlopen maar tegelijk de kernelementen van effectieve samenwerking ondermijnden (Johnson & Johnson, 1984; 2017). Opvallend was dat leerkrachten deze aanpassingen deden tijdens de tweede fase van de taken die volgens jigsawprincipes waren vormgegeven (zie Aronson, 1978). Lesobservaties lieten zien dat leerlingen vanaf dat punt vaak *minder* taakgericht gedrag vertoonden en hun taak minder vaak voltooiden.

In lijn met eerdere studies blijkt het effectief begeleiden van samenwerkend leren binnen een klassikale setting een uitdaging te vormen voor leerkrachten (zie Hacker & Tenent, 2002; Okkinga et al., 2018b; 2021; Spörer et al., 2009) en leerlingen (De Backer et al., 2021; Veldman et al., 2020), temeer omdat er in het reguliere curriculum zo weinig aandacht voor is door de sterke nadruk op individueel vragen beantwoorden bij een tekst (zie *Hoofdstuk 3*). Tegelijk is het hoopgevend dat zowel leerkrachten als leerlingen juist deze samenwerkingstaken positief beoordeelden.

Tot slot kwam naar voren dat er een zekere spanning werd ervaren tussen de autonomie van leerkrachten en de mate waarin lessen 'gescript' waren. De sterk gescripte lesmaterialen zorgden weliswaar voor houvast en vereenvoudigden de lesvoorbereiding, maar de keerzijde ervan was dat sommige leerkrachten aangaven dat ze – met toenemende kennis – de voorschriften soms te strikt vonden. Het blijkt een pittige evenwichtsoefening om enerzijds via lesmaterialen voldoende steun te bieden aan leerkrachten zonder hun professionele autonomie te ondermijnen.

4. Conclusie

In dit proefschrift beschreef ik of wetenschappelijke theorieën over tekststructuur-onderwijs succesvol naar de Nederlandse onderwijspraktijk vertaald konden worden. De meta-analyse (*Hoofdstuk 2*) rapporteerde *wat* voor soort onderwijs in tekststructuur de grootste effecten heeft (inhoud) en *hoe* dit onderwijs het beste kan worden vormgegeven (didactiek). De mixed-methods analyse (*Hoofdstuk 3*) legde een kloof bloot tussen wetenschap en onderwijspraktijk. Dit riep de vraag op hoe tekststructuuronderwijs er dan *wél* uit zou kunnen zien en vormde de opmaat naar het ontwerpgericht onderzoek waarbij de op theorie- en empirie-gebaseerde puzzelstukjes in elkaar werden gepast tot lesmaterialen (*Hoofdstuk 4*). Daarbij bleek de kloof tussen wetenschap en praktijk soms breder dan gedacht. De uiteindelijke lessenserie werd getoetst in een effectonderzoek met twee interventierondes (*Hoofdstuk 5*). De uitkomsten bleken minder eenduidig dan verwacht: in één interventieronde liet een selecte groep leerlingen alle verwachte effecten zien. Deze uitkomst is besproken in het licht van een beperkte toetskwaliteit en implementatieproblemen. De multiple-casestudy (*Hoofdstuk 6*) belichtte implementatieproblemen, vooral op het didactische vlak.

Het huidige onderzoek draagt bij aan theorievorming rondom de bestaande kennis en professionaliseringsbehoefte van leerkrachten en rondom kansrijke didactische aanpakken. Daarnaast zijn er vragen opgeroepen die in vervolgonderzoek aandacht zouden moeten krijgen, bijvoorbeeld over de effecten van specifieke leerkrachtkennis, implementatiekwaliteit en didactische keuzes op de mate van interventiesucces in het leesonderwijs. Bovendien is meer theorievorming nodig over het betrouwbaar meten van hogere-orde-tekstbegrip.

De in dit proefschrift verkregen inzichten zijn van belang voor de doorontwikkeling van evidence-based maar leerkrachtvriendelijke lesmaterialen die voortborduren op de huidige kennisbasis van leerkrachten. Dit kan een belangrijke stap vormen om de kwaliteit van het leesonderwijs structureel te versterken, maar vereist een gezamenlijke inzet van beleidsmakers, onderzoekers, educatieve uitgevers, pabo's, leerkrachten en andere partijen: alleen met gebundelde krachten kunnen we duurzame en solide bruggen bouwen tussen wetenschap en praktijk.

Dankwoord

'We kunnen deze vlieger beter weggooien en naar huis gaan,' zei Pad.
'Pad', zei Kikker, 'we moeten het nog één keer proberen. Zwaai de vlieger
boven je hoofd. Spring erbij op en neer en roep dan: Ga omhoog, vlieger!'

Pad rende door de wei. Hij zwaaide de vlieger boven zijn hoofd. Hij sprong
erbij op en neer. Hij schreeuwde: 'Ga omhoog, vlieger!' De vlieger ging
omhoog. Hij steeg hoger en hoger.

'Hij doet het!' riep Pad.
'Ja natuurlijk,' zei Kikker.
'Toen het niet lukte met rennen,
en toen het niet lukte met rennen en zwaaien,
en toen het niet lukte met rennen, zwaaien en springen,
wist ik dat het moest lukken met rennen, zwaaien, springen en roepen.'

Uit: Alfred Lobel, *Kikker en Pad*

Huub van den Bergh, dank dat je de afgelopen jaren mijn weergaloze promotor wilde zijn. Misschien herinner je je nog dat je ons het boek van *Kikker en Pad* gaf als kraamcadeau. Niet alleen een geweldig voorleesboek, maar vooral ook zo vol met toepasselijke lessen. Met veel geduld en humor hielp je me bij het aanscherpen van mijn bevindingen, mijn methode en analyses. Soms moesten we niet alleen rennen, maar moest er een hele analyse op zijn kop. Net als Kikker denk ik dat je altijd wel wist dat je als onderzoeker soms gelijktijdig moet rennen, zwaaien, springen en roepen om iets voor elkaar te krijgen. Dank voor je geduld, je scherpe blik, je humor, je bemoediging en je prachtige verhalen over nieuwsgierige postbodes, adellijke portretten en (klein)kinderen opvoeden. Je was een allerm minst koele, maar wel zeer wijze Kikker op mijn pad.

Een taart bakken, dat kan ik toch zeker ook wel, dacht Kikker toen hij weer thuis was. Hij gooide alles wat hij vinden kon in een kom en begon te roeren. Zo had hij het Varkentje ook zien doen. Daarna deed hij het in een koekenpan en zette die op het vuur. Ziezo, dacht Kikker, dat wordt smullen. Maar na een tijdje begon het vreselijk te roken en te stinken. De taart was helemaal verbrand.

Uit: Max Velthuis, *Kikker is Kikker*

Jacqueline Evers-Vermeul, dank dat je mijn ongeëvenaarde copromotor was. Jij voorzag me van een prentenboek over een andere Kikker. Eveneens vol met wijze lessen vakdidactiek: zo zie je dat het observerend leren van Kikker toch nog wat te wensen overlaat. Nu is taarten leren bakken nog wel wat anders dan diep tekstbegrip opbouwen, maar het was een mooi avontuur om op dat vlak samen tot nieuwe inzichten en vragen te komen. Dank voor je geduldige feedback als ik weer eens talloze ingrediënten in mijn schrijfsels gooide; volgens mij hebben we er samen best wat van gebakken. Ik heb me altijd verwonderd over je gave om moeiteloos dubbele spaties te spotten. Speciaal voor jou heb ik er eentje toegevoegd in dit dankwoord, maar die heb je ongetwijfeld al gezien. Dank voor je toewijding en je betrokkenheid. Je maakte de reis hiernaartoe tot een feest!

Onmisbaar was de inzet van zoveel leerkrachten en leerlingen die hun ervaringen met mij deelden en me soms urenlang tolereerden als toeschouwer in het klaslokaal. Een eervolle vermelding voor Monique, Bettina, Marjolein en Christie die als ware pioniers een lessenserie ontwierpen om de kloof tussen wetenschap en praktijk een stukje verder te overbruggen, ondanks alle uitdagingen.

Een speciaal woord van dank aan mijn paranimfen en collega's Nina Sangers en Hilde Kooiker-den Boer. Het was fijn om samen te sparren als we door de bomen het bos even niet zagen, om te bomen en te zagen over het lastige iken van testitems. Ik heb mooie herinneringen aan Brighton en het hutje-zonder-hei en de vele kopjes koffie en thee. Dank dat jullie het nooit zat waren om me te horen

klagen over deadlines en flesweigerende baby's. Jullie vriendschap maakte mijn PhD nog vele malen gezelliger.

Dank ook aan mijn leescommissie: wat fijn dat jullie de tijd hebben genomen om dit – best wel dikke – boek kritisch te lezen. Dank aan alle anderen die op welke wijze dan ook hun steentje hebben bijgedragen, in het bijzonder José van der Hoeven, Anja de Wijs, Gert Rijlaarsdam, Kees Vernooy, Rémi Bouwer, Abne Herrebout, and my overseas colleagues, Kay Wijekumar and John Strong who shared their expertise with me. Dank aan alle collega's en mede-uiltjes die de Universiteit Utrecht tot een fijn nest maakten, in het bijzonder mijn fantastische kantoorgenoten (Na, Gijs, Kirsten, Yipu, Loes), mijn fijne onderzoeksassistent Puck Colen en de andere (oud-)L&E-promovendi (Deborah, Frederike, Leslie, Corina, Patrick, Florentine, Abid, Klaske, Esther, en alle anderen).

Lieve familie, vrienden en vriendinnen, papa en mama: dank dat jullie vol geduld en interesse mijn leven en onderzoek gevolgd hebben. Met taartjes en theetjes, zelfgemaakte pindakaas, rode bietjes uit eigen tuin en talloze appjes vol baby- en huisdierspam hebben jullie voor heerlijke afleiding gezorgd. Bijzonder veel dank aan (o)ma Bogaerds die zo trouw elke dinsdag kwam oppassen en óók nog eens de stapels wasgoed stond te strijken met de grootste glimlach; en dank aan wederhelft (o)pa voor de stapels pannenkoeken die de dinsdagen tot een nóg groter feest maakten.

Liefste David, jij bent de wind onder mijn vlieger. Zonder jouw liefdevolle aansporing zou ik niet eens een PhD overwegen. Wat een rijke zegen dat we de afgelopen jaren ook onze lieve dochters Louise, Elodie en Madeleine mochten ontvangen uit Gods hand. Het was uitdaging en pure zegen om een proefschrift te schrijven met schurend zandbakzand en nieuwsgierige kindervingertjes in mijn toetsenbord – en ik had dat voor geen goud willen missen.

Bovenal dank aan God die ons elke dag gedragen heeft.
Want uit Hem en door Hem en tot Hem zijn alle dingen.

About the author

Suzanne Bogaerds-Hazenberg (1990) obtained bachelor degrees in Pedagogical Sciences (cum laude) and Romance Languages and Cultures (summa cum laude), and completed the research master Language and Cognition (summa cum laude) at the University of Groningen. During her studies, Suzanne worked for an educational publisher, and assisted in various research projects on (second) language teaching, children's language acquisition, and education.

The current research was funded by a NWO Research Talent Grant (406-16-052) and was carried out at Utrecht University. Suzanne worked on her PhD on a parttime basis under supervision of prof. Huub van den Bergh and dr. Jacqueline Evers-Vermeul. The main findings were presented at various scientific conferences (SSSR, ST&D, World Literacy Conference, AILA, Anéla, ORD). Suzanne also gave workshops for teachers at various national events (Landelijk Netwerk Taal, ResearchED, Kenniscentrum Begrijpend Lezen). She shared findings from her research with various institutions and policy makers (SLO, Dutch Inspectorate of Education, and the Flemish Department of Education).

Suzanne recently started as a postdoctoral researcher at the Radboud University to examine the implementation quality of a close reading intervention for struggling readers.





Proficient readers often make predictions about the text content based on the text's skeleton or underlying text structure. Readers who recognize the underlying text structure will already have a pretty good image of what kind of information can be expected and how the main ideas will be organized. This dissertation examines whether explicit text structure instruction can foster primary school students' reading comprehension skills. After a meta-analysis on the effects of text structure instruction, and an examination of the current reading comprehension curriculum in the Netherlands, this dissertation reveals through a design-based research, an intervention study, and a multiple-case study both the possibilities of text structure instruction in the Dutch context, as well as the many obstacles that need to be overcome before we can build sustainable bridges between research and practice in the area of reading comprehension instruction.
